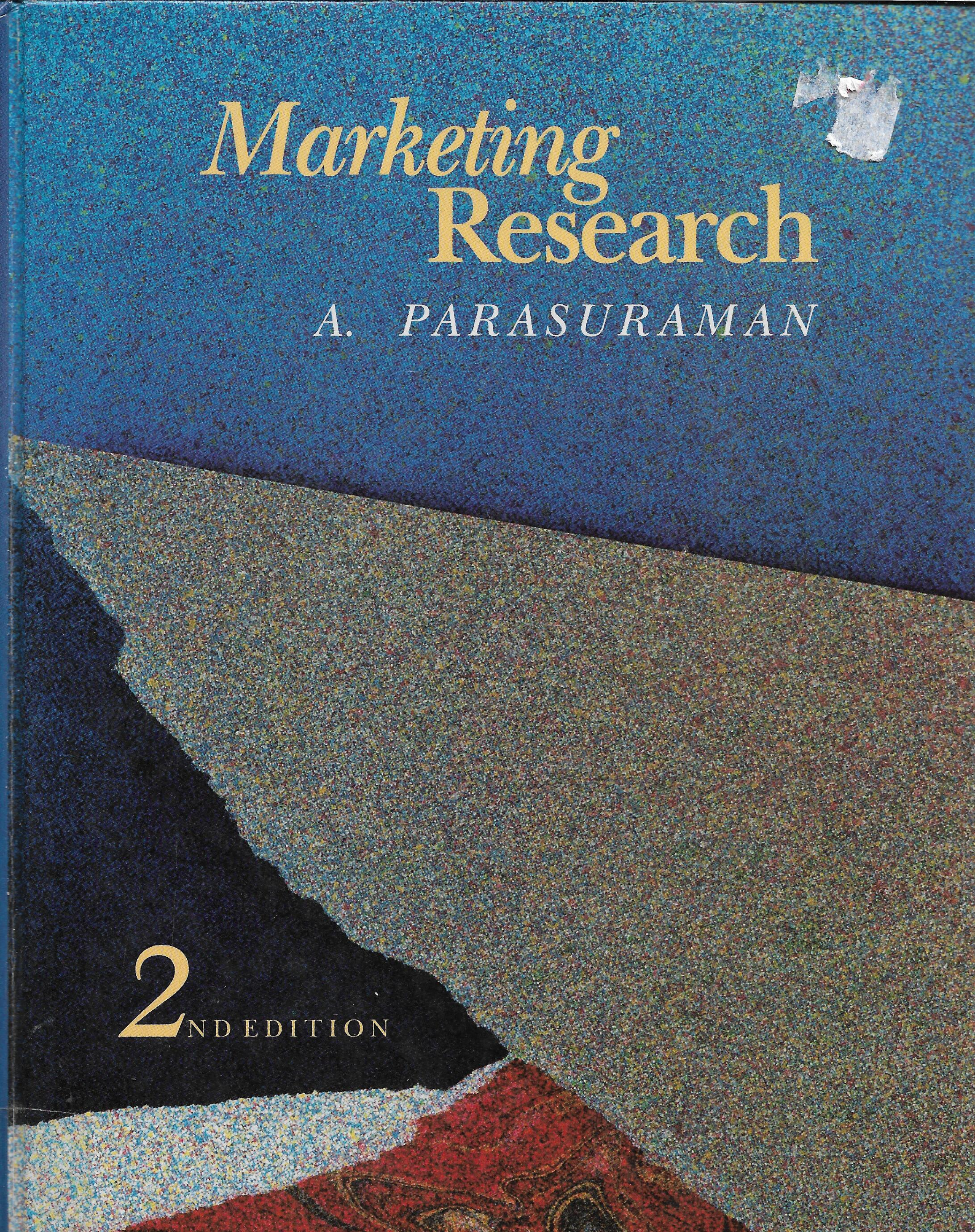


Marketing Research

A. PARASURAMAN

2ND EDITION

To my wife Ranga,
sons Rohith and Vishnu,
and daughter Roopa

Editor-in-Chief: Robert Price Jr.
Production Supervisor: Kazia Navas
Packaging Service: Barbara Gracia, Woodstock Publishers' Service
Production Coordinator: Jason Jordan
Text Designer: Gary Fujiwara
Cover Designer: Marshall Henrichs, Gary Fujiwara
Technical Art Consultant: Joseph Vetere
Illustrator: Scot Graphics
Copyeditor: Diana Gibney
Permissions Editor: Mary Dyer
Manufacturing Supervisor: Roy Logan

Library of Congress Cataloging-in-Publication Data

Parasuraman, A.
Marketing research/A. Parasuraman.—2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0–201–50282–8
1. Marketing research. I. Title.
HF5415.2.P37 1991
658.8′3—dc20 90-35939
CIP

Copyright © 1991 by Addison-Wesley Publishing Company, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ABCDEFGHIJ-HA-94321

Preface

Studying marketing research can be an enlightening, interesting, and rewarding experience, contrary to what many students believe. The content and the format of *Marketing Research*, Second Edition, are designed to aid in offering such an experience, especially to the typical student who approaches the subject apprehensively and reluctantly. Like the first edition, this edition discusses the various marketing research terms, tools, and concepts in an understandable, lively fashion by using numerous examples in every chapter. In addition, it incorporates the impact on marketing research of recent advances in computer and information technologies.

AUDIENCE

This book is intended for students who are taking their first course in marketing research. The breadth and the depth of topics included in the book are carefully chosen to benefit those students who will primarily be potential research users as well as those who wish to consider marketing research as a career. To cater to this dual audience, the book stresses the importance of effective communication between research users and researchers and highlights their respective roles and responsibilities at various stages of a research project.

CONTENT AND ORGANIZATION

The book contains twenty chapters, organized into seven parts. Part One, consisting of two chapters, discusses the nature and scope of marketing research, the various parties involved in conducting marketing research, and the increasing sophistication and capabilities of research-based systems to aid decision making. Part Two, also consisting of two chapters, offers an overview of the various research project stages, the interrelationships among the stages, and the basic research approaches and designs available for conducting marketing research.

Part Three contains five chapters. These chapters describe different types of data and discuss alternative methods for obtaining data. Part Four is made up of two chapters that focus on designing data collection forms and instruments.

Four chapters make up Part Five. The first three chapters cover sampling techniques and issues germane to generalizing results obtained from a sample. The fourth chapter discusses sampling and nonsampling errors that enter the data collection process.

Part Six consists of four chapters and deals with data analysis. The first chapter focuses on preparing the raw data for analysis, conducting certain preliminary analyses, and getting a feeling for the types of additional analyses that may be appropriate. The last three chapters in this part discuss a variety of specific analysis techniques.

Part Seven, containing the final chapter of the book, deals with communicating research findings to decision makers. It offers a number of guidelines for preparing written reports and making oral presentations that succinctly describe a research project and effectively convey its findings to the target audience.

The content and organization of the first edition were influenced by an extensive survey conducted by Addison-Wesley to ascertain the needs and the concerns of college faculty who teach marketing research. This survey produced a wealth of useful advice about what topics and features to include in the book, how much emphasis to place on them, and how best to present them. The second edition, in addition to preserving the strengths of the first edition, incorporates a number of excellent suggestions received through another survey of faculty exposed to the first edition. I sincerely thank those who participated in the surveys. Their collective responses helped me to focus on issues and ideas important for today's marketing research environment.

Colleagues who meticulously reviewed and critiqued parts or all of the book at various stages of its development have been another valuable source of ideas and inspiration for me. The following individuals reviewed drafts of one or both editions of the book: O. Ahtola, University of Denver; William O. Bearden, University of South Carolina; Richard F. Beltramini, Arizona State University; Larry Compeau, Virginia Polytechnic Institute; James D. Culley, DuPont Company; Phillip E. Downs, Florida State University; Robert Eng, Babson College; Raymond P. Fisk, University of Central Florida; Peter J. Gordon, Southeastern Missouri State University; Vince Howe, University of Kentucky; Roy D. Howell II, Texas Tech University; Michael Hyman, University of Houston; Linda F. Jamieson, Northeastern University; Glen R. Jarboe, University of Texas at Arlington; Pamela Kiecker, Texas Tech University; Robert E. Krapfel, Jr., University of Maryland; Nicholas Nugent, New Hampshire College; Susan M. Petroshius, Bowling Green State University; William Redmond, Bowling Green State University; Coskun Samli, University of North Florida; William J. Sauer, University of Pittsburgh; Alan G. Sawyer, University of Florida; Robert Schaffer, California State Polytechnic University—Pomona; David Snepenger, Montana State University; John H. Summey, Southern Illinois University at Carbondale; David J. Urban, Georgia State University; Ganesan Visvabharathy, Northern Illinois University; George M. Zinkhan, University of Houston. I am very grateful to these reviewers for their constructive criticisms, insightful suggestions, and words of encouragement.

FEATURES NEW TO THIS EDITION

This edition has a number of new features that should appeal to instructors and students alike.

1. A new chapter titled "Marketing Research and Decision Support Systems" replaces Chapter 2 of the previous edition. (The material in Chapter 2 of the first edition has been condensed and integrated into a revised Chapter 1.) This unique chapter includes concepts and applications that are at the cutting edge of recent developments in gathering and effectively using marketing information. It traces the evolution of present-day information systems—from *ad hoc* marketing research projects to marketing information systems to marketing decision support systems to expert systems—and discusses the role of marketing research in such systems.
2. This edition includes two new cases—"National Insurance Company" (Case 6.5) and 'Fantastic Car Wash" (Case 6.6)—with data sets. The data sets are derived from real-life surveys and include a variety of metric and nonmetric data. The National Insurance Company data set has 285 observations on 37 variables, and the Fantastic Car Wash data set has 256 observations on 24 variables. Both data sets are on a floppy disk packaged with the Instructor's Manual.
3. Every chapter contains several "Research in Use" boxed examples that illustrate real-world applications of concepts or techniques covered in the chapter. These examples are integrated well with the main chapter material.
4. Every chapter begins with a set of learning objectives built around all major topics or concepts covered in the chapter.
5. Key terms in each chapter are highlighted in the text, defined in the margins, and listed at the end of the chapter for quick review. The book also has a comprehensive glossary containing an alphabetical listing of all key terms with their definitions.

6. This edition is printed in two-color format for enhanced graphics and visual appeal.
7. Chapters dealing with data analysis techniques (Chapters 17–19) present illustrations of actual computer output *and discuss* how to interpret the output.
8. The first chapter on sampling (Chapter 12) and the chapters on data analysis techniques (Chapters 17–19) provide introductory overviews of the methods to be covered. These overviews contain flowcharts or tables that serve as "road maps" for students.
9. The chapter on presentation of research results (Chapter 20) has an expanded discussion of report writing and the use of computer graphics to effectively represent information in one-way tables and cross-tabulations.
10. The material in a number of chapters, especially Chapters 1, 4, 11, 16, and 17, is now streamlined and/or rearranged for better clarity.

OTHER SPECIAL FEATURES

This edition of the book preserves, and in many instances enhances, the following distinct characteristics of the previous edition.

1. Every chapter in the book has numerous, carefully chosen examples that are well integrated with the textual material. Every key concept, principle, or technique is illustrated with one or more examples.
2. The book has a readable and accessible writing style. The key ideas in each chapter are logically developed and organized to enable students to grasp them clearly and quickly. Charts, diagrams, and other graphical illustrations are used liberally to supplement the textual discussion.
3. The book offers useful frameworks and step-by-step approaches for understanding and applying effectively the more technical aspects of marketing research (e.g., the advanced probability-sampling techniques covered in Chapter 13 and the hypothesis-testing procedures covered in Chapter 17).
4. The book provides extensive coverage of the relatively "soft" marketing research procedures that are frequently used in practice but are treated only superficially in many textbooks. For example, this book has a separate chapter on qualitative research (Chapter 7) and one on nonprobability sampling (Chapter 14).
5. The questions at the end of chapters discussing some form of data manipulation (e.g., Chapters 12 and 18) include a number of problems that require students to apply the procedures or techniques discussed.
6. Appendix A at the end of the book contains a real-life data set derived from consumer panel data provided by NPD Research, Inc., of New

York. This appendix provides a detailed description of the data set (including labels for and definitions of the variables included) and contains 25 student exercises grouped according to the analysis chapters to which they relate. The data set contains cross-sectional as well as longitudinal data, and it includes variables reflecting metric as well as nonmetric measurement levels. Data are available for 100 families on 58 variables. This data set (along with the data sets for Cases 6.5 and 6.6) is on a floppy disk packaged with the Instructor's Manual.

7. The book includes 24 cases (most of them built around true scenarios) that cover a variety of products, services, and situations.

FLEXIBILITY FOR INSTRUCTORS

The content and the organization of this book are designed to offer maximum flexibility for instructors in terms of the teaching approaches they want to use and the relative emphases they wish to place on different topics. The simple writing style and the liberal use of examples should minimize the amount of class time necessary to merely clarify concepts covered in the book. Instructors will therefore have more time for further developing key topics they wish to emphasize and for experiential activities such as case discussions, computer-assisted exercises, and student projects.

Instructors who wish to place more emphasis on student projects and less on the textual material itself can do so by omitting some or all sections of certain chapters. For example, instructors who do not want to cover experimentation in detail can omit Chapters 8 and 9 (students can still gain insights about the basic ideas underlying experimentation from the material covered in Chapter 4). Similarly, in the area of data analysis, instructors who do not want to go much beyond tabulation procedures and computation of summary statistics can cover Chapter 16 in detail but skip some or all of Chapters 17, 18, and 19. In short, while this book covers a wide range of topics, its content and organization are flexible enough to help instructors choose the breadth and depth of topical coverage they desire.

INSTRUCTIONAL SUPPORT

This edition of *Marketing Research* has a greatly expanded and strengthened package of support materials.

Instructor's Manual with Transparency Masters. Coäuthored with Daryl McKee of Louisiana State University, the instructor's manual contains a variety of chapter support materials. These include chapter outlines, discussion examples not in the book, tips for covering the chapter material, student miniprojects that can be used in class or as homework assignments, answers to end-of-chapter questions and problems, and annotated additional readings. The instructor's manual also provides sample course outlines, solutions to all cases and data analysis exercises, and over 100 transparency masters.

Test Bank. Developed by Thomas J. Quirk of Webster University, the test bank contains about 1400 true/false and multiple-choice questions, several short-answer and essay questions for each chapter, and computational/problem-solving questions for Chapters 12, 13, 17, and 18. It also includes a separate page for each chapter wherein the true/false and multiple-choice questions are coded by difficulty level (low, medium, or high) and are linked to the text by page number.

Computerized Test Bank. All true/false and multiple-choice questions are also available on a floppy disk for IBM and IBM-compatible personal computers. This computerized test bank allows instructors to edit, add, or delete questions while generating tests.

Data Disk. The NPD data set and the two case data sets are available on a floppy disk which is packaged with the Instructor's Manual. These data sets can be used with any statistical software packages, such as SPSS-X, SAS, Minitab, etc.

ACKNOWLEDGMENTS

In addition to the survey respondents and reviewers mentioned earlier, a number of other individuals and organizations contributed significantly to the development of this textbook. My sincere appreciation goes to the numerous organizations and executives who generously provided material for the cases, exercises, and illustrations in this book. I also thank my graduate assistants—in particular, Irfan Ahmed, Wenda Leong, Chris Wilkes, and Bryan Wilson—and my secretary, Glenda Bessler, who successfully tackled the many chores associated with preparing this edition. The secretarial staff at Texas A&M's marketing department provided me with timely typing assistance for which I am grateful. I also owe a sincere debt of gratitude to Mary Fisher, Marilyn Freedman, Maribeth Jones, Christine O'Brien, and other members of Addison-Wesley's editorial staff for their continuous support. Several individuals whose expert assistance was crucial in transforming my manuscript into this textbook include Diana Gibney, my copyeditor, Barbara Gracia of Woodstock Publishers' Services, and Kazia Navas of Addison-Wesley. Finally, words alone are insufficient to thank my wife and children for their generous understanding and support throughout this project.

College Station, Texas **A. P.**

Contents

Part One
Introduction

1 Nature and Scope of Marketing Research 4
- Chapter Objectives 4
- Definition of Marketing Research 5
- Marketing Research's Role in Decision Making 12
- Providers of Marketing Research Services 15
- Marketing Manager–Marketing Researcher Interface 19
- Ethics in Marketing Research 23
- Summary 31
- Key Terms 32
- Questions 32
- Appendix to Chapter 1: Guidelines for Using External Research Providers 35

2 Marketing Research and Decision Support Systems 40
- Chapter Objectives 40
- Evolution of Systems for Supporting Decision Making 42
- Marketing Information System 44
- Marketing Decision Support System 49
- Expert System 57
- Summary 60
- Key Terms 61
- Questions 61
- Appendix to Chapter 2: Careers in Marketing Research 63

Cases for Part One

- Case 1.1: Action Market Research Services, Inc. 72
- Case 1.2: Modern Silo Systems 75
- Case 1.3: Walker Research, Inc. 78

Part Two
Key Aspects of the Marketing Research Process

3 Steps in a Marketing Research Project 84

- Chapter Objectives 84
- Deciding Whether a Marketing Research Project Is Worth It 85
- Defining the Project's Purpose 92
- Identifying Data Needs 96
- Identifying Data Sources 97
- Choosing an Appropriate Research Design 98
- Developing Data Collection Forms 100
- Designing the Sample 100
- Collecting the Data 100
- Analyzing Data and Interpreting Results 101
- A Real-Life Illustration 102
- Summary 109
- Key Terms 110
- Questions 110
- Appendix to Chapter 3: Bayesian Analysis 113

4 Types of Marketing Research 126

- Chapter Objectives 126
- Exploratory Versus Conclusive Research 128
- Conducting Exploratory Research 130
- Descriptive Versus Experimental Research 135
- Conducting Descriptive Research 137
- Conducting Experimental Research 146
- Determining Which Type of Research to Conduct 148
- Summary 150
- Key Terms 151
- Questions 151

Cases for Part Two 153

- Case 2.1: Bettcher Industries, Inc. 154
- Case 2.2: Fast-Test: To Use or Not to Use? 158
- Case 2.3: Superior Landscape Company 163

Part Three
Data Collection Procedures

5 Secondary Data 168

- Chapter Objectives 168
- Advantages of Secondary Data 170
- Limitations of Secondary Data 172
- Sources and Types of Secondary Data 180
- Managing Secondary Data 191
- Summary 192

Key Terms 193
Questions 193
Appendix to Chapter 5: External Secondary-Data Sources 195

6 Overview of Primary-Data Collection Methods 206
Chapter Objectives 206
Variety of Data Collection Methods 207
Questioning Versus Observation 209
Questionnaire Format 216
Questionnaire Administration Methods 229
Types of Observation Techniques 240
Summary 247
Key Terms 248
Questions 249

7 Qualitative Research 250
Chapter Objectives 250
What Is Qualitative Research? 251
Qualitative Versus Quantitative Research 252
Importance of Qualitative Research 252
Focus Group Interviews 255
Other Qualitative-Research Techniques 273
Summary 276
Key Terms 278
Questions 278

8 Introduction to Experimentation in Marketing 280
Chapter Objectives 280
Descriptive Versus Experimental Research 282
Conditions for Inferring Causality 283
Laboratory Versus Field Experiments 285
Deciding on Which Type of Experiment to Use 289
Threats to Internal Validity 295
Threats to External Validity 302
Concluding Remarks on Validity Threats 305
Summary 306
Key Terms 307
Questions 308

9 Selected Experimental Designs 310
Chapter Objectives 310
Pre-experimental Designs 312
True Experimental Designs 317
Quasi-Experimental Designs 328
Other Experimental Designs 330
Summary 338

Key Terms 340
Questions 340

Cases for Part Three 343
Case 3.1: WTTW–Chicago 349
Case 3.2: Hamilton Bank (A) 349
Case 3.3: Sid's Sandwich Shops 351
Case 3.4: Hershey Chocolate Company (A) 353
Case 3.5: Professional Ski Supply Company 357

Part Four Designing Forms and Scales for Collecting Data

10 Designing Data Collection Instruments 362
Chapter Objectives 362
Questionnaire Design 363
Complexity of Questionnaire Design 364
Questionnaire's Impact on Data Accuracy 365
Questionnaire Design Elements 367
Question Form 367
Question Relevance and Wording 376
Sequencing of Questions 387
Questionnaire Appearance and Layout 394
Pretesting 396
Questionnaires for Computerized Interviewing 397
Designing Cover Letters for Mail Questionnaires 399
Openers for Personal and Telephone Interviews 399
Designing Observation Forms 401
Summary 403
Key Terms 404
Questions 405

11 Types of Scales and Attitude Measurement 406
Chapter Objectives 406
Measurement Levels 408
Classes of Variables 414
Attitude Scaling 417
Self-Report Measurements of Attitudes 420
Commonly Used Multiple-Item Scales 433
Strengths of Multiple-Item Scales 441
Summary 445
Key Terms 447
Questions 447

Cases for Part Four 449
Case 4.1: People's National Bank 450
Case 4.2: Bankers Life and Casualty Company 452
Case 4.3: Hershey Chocolate Company (B) 461

Part Five Designing the Sample and Collecting the Data

12 Introduction to Sampling 472

- *Chapter Objectives 472*
- *What Is Sampling? 473*
- *Sampling Versus Census Studies 474*
- *Conditions for Doing Census Studies in Marketing 477*
- *Probability Versus Nonprobability Sampling 478*
- *Definition of Simple Random Sampling 480*
- *Choosing a Simple Random Sample in Practice 482*
- *Sampling Error and Sampling Distribution 485*
- *Confidence-Interval Estimation 490*
- *Determining Sample Size 495*
- *Limitations of Simple Random Sampling 502*
- *Summary 502*
- *Key Terms 504*
- *Questions 504*

13 Advanced Probability-Sampling Methods 508

- *Chapter Objectives 508*
- *Stratified Random Sampling 509*
- *Proportionate Stratified Random Sampling 510*
- *Disproportionate Stratified Random Sampling 520*
- *Cluster Sampling 522*
- *Simple Cluster Sampling 523*
- *Area Sampling 525*
- *Systematic Sampling 529*
- *Comparison of the Various Methods 533*
- *Summary 535*
- *Key Terms 536*
- *Questions 536*

14 Nonprobability-Sampling Methods 540

- *Chapter Objectives 540*
- *Convenience Sampling 541*
- *Judgment Sampling 544*
- *The Use of Convenience and Judgment Samples 546*
- *Quota Sampling 548*
- *Probability-Versus Nonprobability-Sampling Methods 553*
- *Summary 559*
- *Key Terms 560*
- *Questions 560*

15 Data Collection: Fieldwork and Potential Errors 562

- *Chapter Objectives 562*
- *Sampling Versus Nonsampling Errors 564*
- *Types of Nonsampling Errors 567*

Controlling Nonsampling Errors 575
Summary 587
Key Terms 589
Questions 589

Cases for Part Five 591
Case 5.1: Dexter's Department Stores 592
Case 5.2: Northwest Bank of Woodcreek (A) 595
Case 5.3: Phoenix Newspapers, Inc. 599
Case 5.4: McCaw Cablevision 602

Part Six
Data Analysis

16 Quality Control and Initial Analysis of Data 606
Chapter Objectives 606
Editing 608
Coding 614
Preliminary Data Analysis 625
Summary 639
Key Terms 640
Questions 640

17 Hypothesis Testing and Its Role in Data Analysis 642
Chapter Objectives 642
Descriptive Versus Inferential Analysis 643
Overview of Hypothesis Testing 643
Role of Hypothesis Testing in Data Analysis 654
Specific Hypothesis Tests 655
Summary 677
Key Terms 678
Questions 679
Appendix to Chapter 17: Relationship of Power to Significance Level and Population Parameter Value 682

18 Techniques for Examining Associations: Chi-Square, Correlation, and Regression 686
Chapter Objectives 686
Chi-Square Contingency Test 687
Spearman Correlation Coefficient 696
Pearson Correlation Coefficient 699
Simple Regression Analysis 704
Deriving a Regression Equation 705
Evaluating the Regression Equation 707
Practical Applications of Regression Equations 713
Precautions in Using Regression Analysis 714
Multiple Regression Analysis 718
Multicollinearity's Implications 720
Summary 723

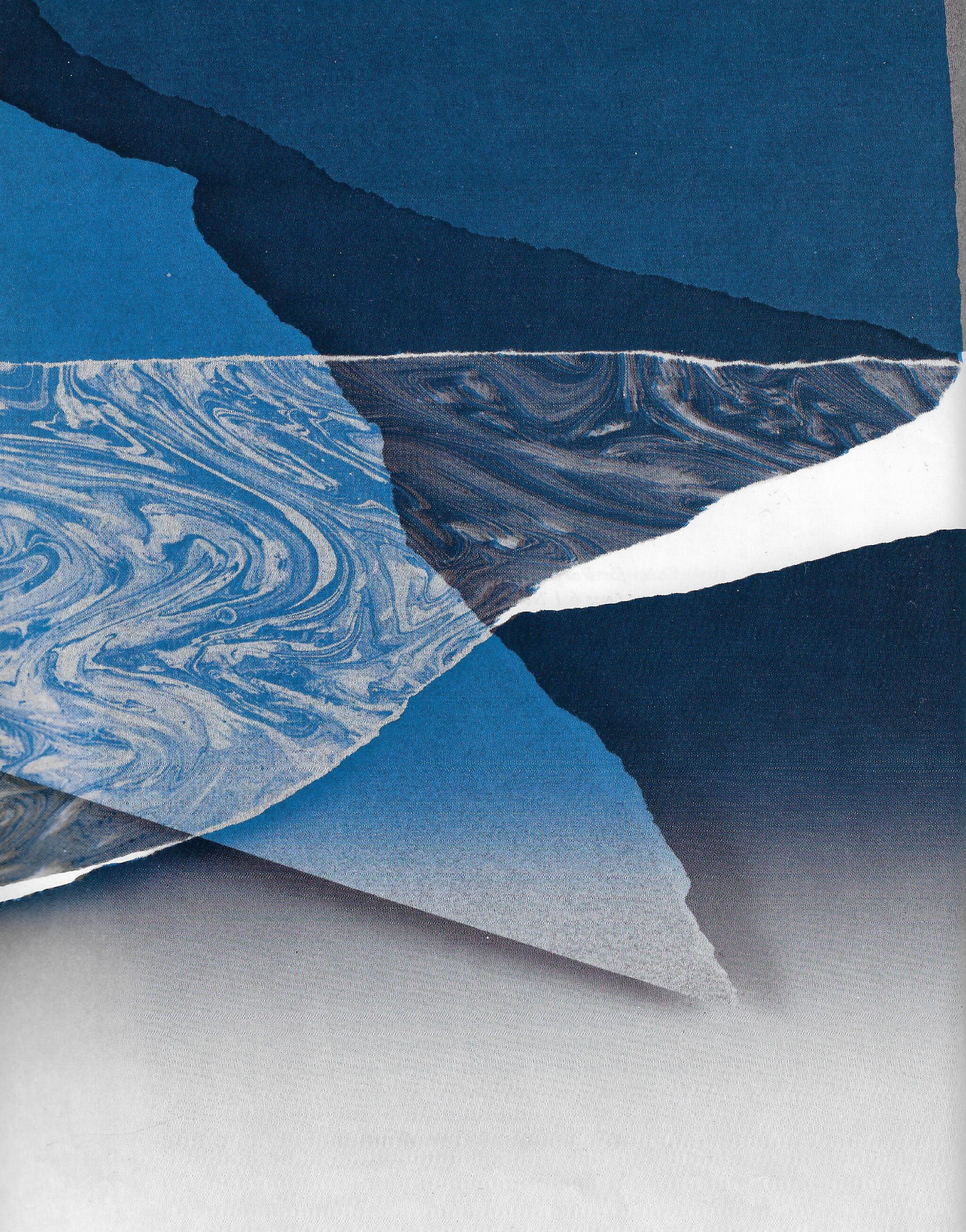
Key Terms 725
Questions 725
Appendix to Chapter 18: Some Technical Details about Correlation and Regression Analysis 729

19 Overview of Other Multivariate Techniques 734
Chapter Objectives 734
Analysis of Variance 735
Advanced ANOVA Procedures 741
Discriminant Analysis 748
Factor Analysis 757
Cluster Analysis 765
Multidimensional Scaling 767
Conjoint Analysis 771
Summary 775
Key Terms 776
Questions 777
Appendix to Chapter 19: ANOVA and Dummy-Variable Regression 779

Cases for Part Six 785
Case 6.1 Northwest Bank of Woodcreek (B) 786
Case 6.2 ITT's Image Advertising Campaign 789
Case 6.3 Ruby Restaurant 792
Case 6.4 Hamilton Bank (B) 796
Case 6.5 National Insurance Company 799
Case 6.6 Fantastic Car Wash 806

Part Seven
Communicating with Research Users

20 Presentation of Research Results 814
Chapter Objectives 814
Importance of Understanding the Audience 816
Preparing Effective Written Reports 816
Graphical Illustrations 829
Oral Presentations 841
Summary 843
Key Terms 844
Questions 844
Appendices 847
Glossary 872
Index 887



PART ONE

Introduction

CHAPTER 1 *Nature and Scope of Marketing Research*

CHAPTER 2 *Marketing Research and Decision Support Systems*

CASES FOR PART ONE

CHAPTER 1

Nature & Scope of Marketing Research

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Define marketing research, and describe its key features.*
- *Illustrate marketing research's role in the three major components of marketing decision making.*
- *Distinguish between in-house marketing research and external providers of marketing research.*
- *State the sources of potential friction between marketing managers and marketing researchers, and indicate how such friction can be minimized.*
- *Identify the five parties that are associated with ethical issues in marketing research.*
- *Explain whether or not a code of marketing research ethics can provide clear answers to specific ethical questions.*

DEFINITION OF MARKETING RESEARCH

When beginning business students are asked to define marketing research, they usually come up with the following types of answers:

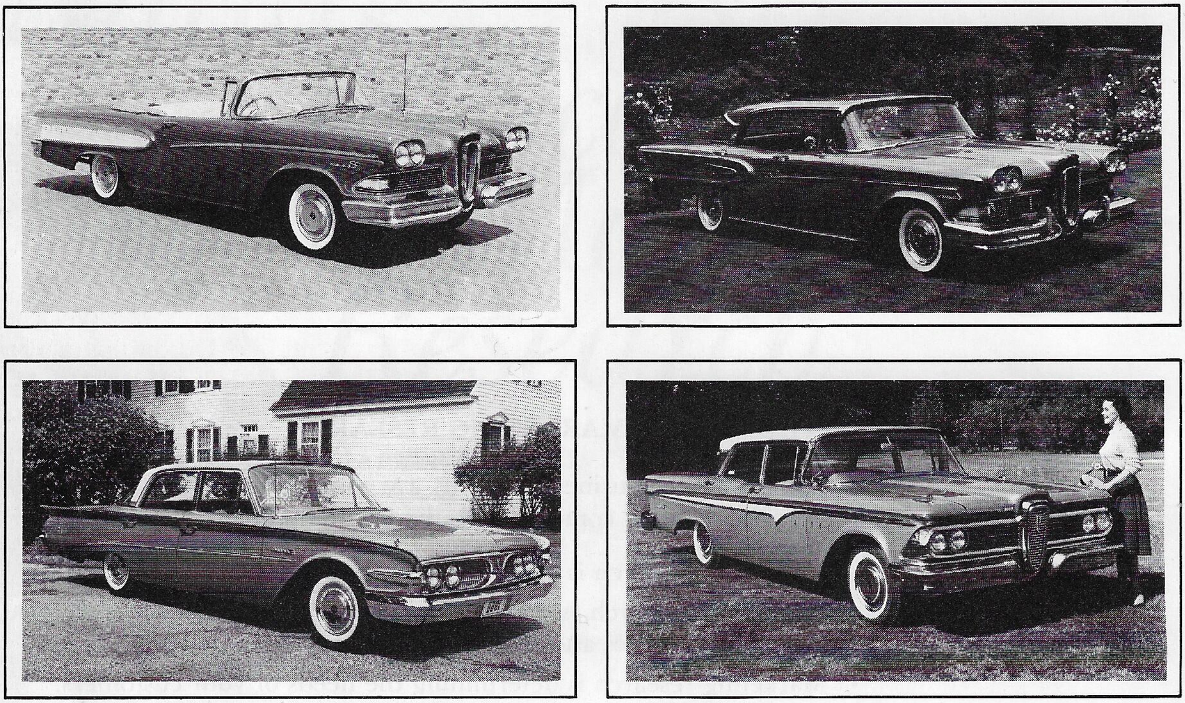
- Marketing research is getting data from your markets.
- Marketing research is a set of quantitative techniques to help you increase your sales and profits.
- Marketing research is determining the needs of your customers.
- Marketing research is gathering sales and market share data of your competitors.
- Marketing research is testing your products in the marketplace.
- Marketing research is estimating the potential sales of your product.

Although these "definitions" describe marketing research, they do not adequately reflect the full scope of it. They are merely examples of potential *applications* of marketing research rather than sound definitions of it. A formal and complete definition of marketing research is given next.

Definition. ***Marketing research*** is a set of techniques and principles for systematically collecting, recording, analyzing, and interpreting data that can aid decision makers who are involved with marketing goods, services, or ideas.

One must carefully consider several key aspects implied by this definition to avoid misconceptions about marketing research and to fully appreciate the nature and scope of it. These aspects are discussed next.

Exhibit 1.1 *Several Models of the Ill-Fated Edsel*



Source: Courtesy of News Department, Ford Motor Company, Dearborn, Mich. 48121.

Marketing Research Involves Qualitative Principles

Marketing research is more than a set of quantitative techniques. It also involves many *principles,* or *guidelines,* for generating managerially useful information. As we will see in the following example, as well as in other sections of the book, these principles are quite simple and, perhaps, intuitive. They are nevertheless crucial to the proper, and potentially successful, application of marketing research. A key reason for the failure of the Edsel automobile, now considered a collector's item, illustrates this point (Exhibit 1.1 shows four Edsel models).

Example. The Edsel, introduced by the Ford Motor Company in 1957, was perhaps one of the most publicized and most costly product failures ever. Ford, which had high sales expectations for the Edsel, lost over $200 million before it was finally withdrawn from the market. Why did the Edsel fail? Contrary to what one might think, the Edsel's failure was not due to a lack of marketing research prior to its introduction. Indeed, Ford did extensive marketing research involving several consumer surveys and sophisticated techniques (such as word association tests) in order to (1) determine the images of existing automobiles with which the Edsel would have to compete and (2) select a suitable name for the automobile.

Unfortunately, the decision makers ignored some of the basic principles of marketing research. First, the image studies and consumer surveys were begun almost 10 years before the Edsel was introduced. Thus the *timeliness and relevance* of the data generated from those studies were questionable. The managers were apparently unaware of a simple, but important, fact about marketing research: *Marketing research can lead to erroneous decisions if it is not done on a timely basis.*

Second, the image studies generated a lot of data about how consumers perceived the automobiles that were already on the market. However, they did not identify the *physical and style features* that would have to be incorporated in the new automobile to make it uniquely appealing to potential customers. Thus the research conducted was not comprehensive enough, apparently because of inadequate consideration of what the research purpose should be and what data should be gathered. *Careful and clear definition of research objectives is a key requirement for beneficial marketing research outcomes.*

Third, the marketing research that was done to select a name for the automobile did *not* recommend the name Edsel. Apparently, management undertook the research merely to see whether the results would support a decision that had already been made. Management thus violated another simple principle: *Marketing research must not be done when potential research users have already made up their minds.*[1]

The resources invested in conducting marketing research for the Edsel were obviously not very productive—*not* because of a failure to employ sophisticated research techniques but, rather, because of inadequate attention to factors such as the timeliness of data, appropriateness of research objectives, and open-mindedness of decision makers. Principles like the ones just illustrated are as crucial to the effectiveness of marketing research as quantitative tools and techniques are.

Marketing Research Involves Analysis and Interpretation

A common misconception about marketing research is that it is merely a set of tools for data collection and recording. Such a misconception seriously underestimates the scope and usefulness of marketing research. *Analysis and interpretation* of data are also well within its domain, as the next example shows.

---

[1] Adapted from Robert F. Hartley, "The Edsel: Marketing Planning and Research Gone Awry," *Marketing Mistakes* (Columbus, Ohio: Grid, 1981), pp. 115–127.

Example. The AmEx Company, which issues the well-known American Express card, undertook some research that resulted in substantial increases in the productivity and profitability of its operations. The research consisted of various *analyses* of data that it collected on a routine basis. For example, analysis of data available from past credit card applications revealed that AmEx took an average of 35 days to process applications for personal charge cards. In contrast, an analysis of records of customer inquiries showed that applicants got impatient if they did not obtain a charge card within three weeks after they had made their applications. As a result, AmEx set a deadline of no more than two weeks to process credit card applications.

Similar analysis of data, related to requests from customers for replacement of lost or stolen cards, revealed that AmEx took two or more weeks to send replacement cards. Furthermore, an analysis of the data related to the dollar amounts charged by customers through their American Express cards showed that AmEx was losing about $2.70 in charge volume for every day a customer was without a card. On the basis of these results, AmEx started replacing cards in as little as two days. Speedy issuance of new cards and replacement cards resulted in additional profits of $1.4 million per year for the firm.[2]

As the example illustrates, if marketing research involves no more than the collection and recording of data, it is highly unlikely to be useful for decision-making purposes. After all, AmEx had been routinely gathering and recording the kinds of data mentioned in the example. However, not until the data were analyzed and the analyses interpreted did the firm reap any tangible benefits from the research. Hence a set of data, by itself, cannot be viewed as useful ***information.*** Analysis and interpretation of data are essential if marketing research is to be truly beneficial to decision makers.

Information *The product of analysis and interpretation of data.*

Marketing Research Does Not Replace Decision Making

Marketing research, *by itself,* cannot be expected to provide sound marketing decisions. The usefulness of marketing research is affected by the ability of potential research users to determine the nature of the problem, the extent to which marketing research can assist in solving the problem, how much money should be allocated to it, what course of action should be taken after observing the results of the research, and so on. For instance, in the Edsel automobile example certain research-related decisions made by the managers—namely, the timing of the automobile image studies and the decision to ignore the results of the name selection research—contributed to the failure of marketing research. In short, although marketing research can aid and influence decision making, the effectiveness of marketing research is itself affected by the decision makers. There is an interactive, or mutual, relationship between marketing research and decision making; therefore we cannot meaningfully view the former as a *replacement* for the latter.

---

[2] "Boosting Productivity at American Express," *Business Week,* October 5, 1981, p. 62.

Another reason marketing research is an aid to decision makers, rather than a replacement for their decision-making process, is that final decisions are influenced not only by research results but also by other factors. A variety of internal factors (e.g., resource constraints, corporate goals) and external factors (e.g., competitors' activities, legal constraints) must be considered in marketing decisions. Therefore decision makers, who should be more knowledgeable about such factors than researchers, have the ultimate responsibility for making decisions and bearing the risk associated with those decisions. The following example illustrates this point.

> **Example.** In the early eighties Nissan Motor Company of Japan, marketer of a variety of products under different brand names in 130 countries, including the well-known Datsun automobile in the United States, started replacing its various brand names with the single brand name Nissan. According to marketing research studies conducted in the United States, the Datsun name had an 85% recognition rate among consumers, while the Nissan name had only a 10% to 15% recognition rate. Furthermore, informal surveys of automobile dealers showed only marginal support for the switch in brand names. Was Nissan Motor Company foolish to go ahead with the name change, considering what the marketing research results seemed to suggest? Perhaps so; perhaps not. A major assumption made by the Nissan management in deciding on the name change was that a single "family" brand name for all its products around the world would ultimately strengthen its corporate identity and improve its marketing efficiency.[3]

This example does not imply that Nissan necessarily ignored the marketing research results. Perhaps the finding that 10% to 15% of U.S. consumers were aware of the Nissan name, although not an overwhelming percentage, might have encouraged management to view *Nissan* as having good recognition potential in the United States if it were adequately promoted. Moreover, the results might have been responsible for the company's cautious strategy of introducing the name change gradually by promoting both the Datsun and the Nissan name during the transition period. That cautious strategy was apparently quite effective, as reflected by the widespread awareness and acceptance the Nissan name currently enjoys. In any event, the point here is that marketing research is only an input into, rather than a replacement for, managerial decision making. The burden of making decisions and assuming the associated risks must be shouldered by management.

---

[3] "A Worldwide Brand for Nissan," *Business Week,* August 24, 1981, p. 80.

Exhibit 1.2 *Marketing Research's Relationship to Decision-Making Activity*

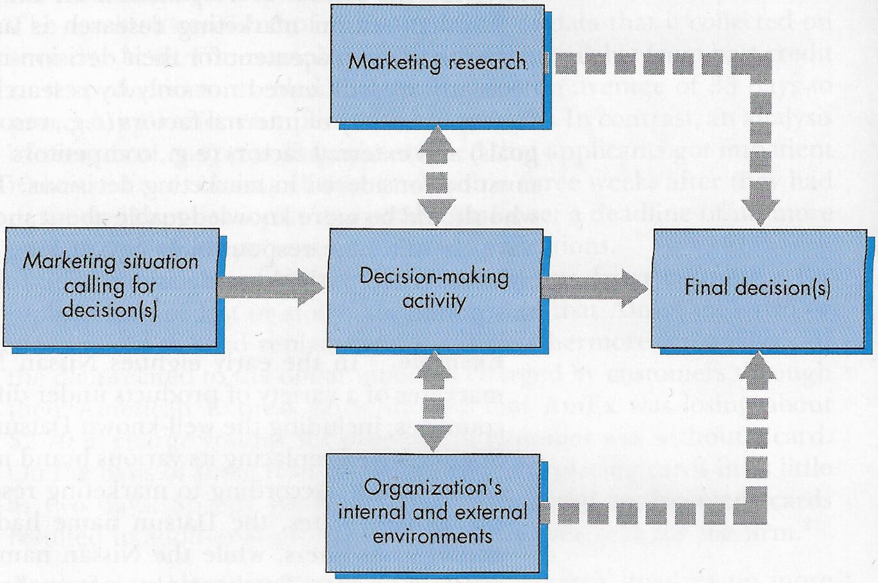

Marketing research
Marketing situation calling for decision(s)
Decision-making activity
Final decision(s)
Organization's internal and external environments

Exhibit 1.2 graphically portrays the relationship of marketing research to decision-making activity. Notice the double-headed arrow showing that marketing research influences and *is influenced by* decision-making activity within an organization. Likewise, there is an interactive relationship between an organization's environment and its decision-making activity. The ultimate responsibility for decisions rests with decision makers, not researchers, as signified by the solid arrow leading from the "Decision-making activity" box to the "Final decision(s)" box. Final decisions may be influenced by, but are not *made* by, those responsible for marketing research.

Marketing Research Has a Variety of Applications

Last, but not least, there are many potential applications for marketing research principles and techniques. In addition to being of value in developing strategies for marketing goods (e.g., canned peaches or chemical fertilizers) and services (e.g., carpet-cleaning services or consulting services), marketing research can be of help in successfully marketing *ideas*. The following example illustrates this point.

Example. Claritas Corporation is a firm that sells information about the profiles and life-styles of Americans on a zip code-by-zip code basis. Jonathan Robbins, founder of Claritas, developed this information base by meticulously combining demographic data compiled by the U.S. Census Bureau and life-style/attitudinal data generated through various regional surveys. The resulting information base contains descriptive profiles of residents in each of the 36,000 five-digit zip code areas in the United States. This information base, nicknamed PRIZM (Potential Rating Index by Zip Market), has found a variety of applications in effectively marketing products *as well as ideas.* For example, several political organizations are clients of Claritas. They use the PRIZM information to pinpoint neighborhoods in which their viewpoints (ideas) are likely to be most (or least) well received. In this way they can effectively allocate their campaign resources to maximize popular support for their viewpoints. As another example, labor unions in Missouri used the PRIZM information base in a successful effort to kill an antiunion right-to-work proposal in a popular vote. The labor leaders identified prolabor neighborhoods with the aid of the PRIZM information base and targeted these neighborhoods for serious campaigning and organizing of voter registration drives.[4]

There are many other such nonbusiness or nonmarketing applications for marketing research. Organizations such as the United Way, American Red Cross, and Campus Crusade for Christ can and do conduct marketing research to aid them in developing strategies for effectively selling their causes to the public and for maximizing the public's support for their programs. Marketing research is also being used increasingly by political candidates who want to run their campaigns more effectively. One of the most extensive applications of marketing research for nonbusiness purposes was made during the 1980 presidential campaign by candidate Ronald Reagan.[5] Yet another novel application of marketing research is the use of customer surveys to determine the value to society of such things as clean air, unspoiled wildernesses, and endangered species.[6] The client lists of large marketing research firms in the United States include many organizations that are nonbusiness. For instance, the client list of Opinion Research Corporation includes the American Economic Foundation, Columbia University, National Education Association, National Society of Professional Engineers, U.S. Department of Transportation, and Veterans Administration.[7]

---

[4] Karen Rypka, "Zipping Up America: Computer Matches Mail Codes to Residents' Lifestyles," *Dallas Morning News,* Section C, April 12, 1982, p. 1C.

[5] For an in-depth discussion of the use of marketing research in Reagan's campaign, see J. J. Honomichl, "The Marketing of a Candidate," *Advertising Age,* December 15, 1980, p. 3.

[6] "How Much Is a Sea Otter Worth?" *Business Week,* August 21, 1989, pp. 59–62.

[7] Names obtained from a client list provided by Opinion Research Corporation, Princeton, N.J.

In summary, marketing research is a versatile set of tools, consisting of qualitative and quantitative principles and procedures that can aid decision makers in a variety of settings. It involves more than merely collecting and storing data or statistics, and its applications go beyond traditional marketing situations—that is, situations involving the marketing of goods or services.

MARKETING RESEARCH'S ROLE IN DECISION MAKING

In any organization a multitude of decisions have to be made; however, our primary focus here will be on the role of marketing research in *marketing* decisions—that is, decisions that are directly related to the marketing of goods, services, or ideas. Specifically, we will examine marketing research's contribution to three major components of marketing decision making: identifying marketing opportunities and constraints, developing and implementing marketing plans, and evaluating the effectiveness of marketing plans.

Identifying Marketing Opportunities and Constraints

Identifying marketing opportunities and constraints is a logical starting point for developing marketing strategies. In particular, firms that are considering new-product introductions or entry into new market areas with existing products will benefit from information about issues such as these: Who are our potential competitors and how strong are their market positions? How satisfied are consumers with current offerings on the market? Are there any unmet consumer needs? How is our offering likely to be perceived by consumers relative to competitors' offerings? Marketing research can help answer these and other questions of interest to marketers.

More and more firms, particularly the larger ones, are using marketing research routinely to generate information relevant for developing effective strategies. For instance, Bram Bluestein, senior vice president of strategic planning at Montgomery Ward, describes the role of marketing research in his company as follows:

> [Marketing research has] almost become a way of life now. Everyone expects that we will go out and do some consumer research as part of researching the business. We typically don't undertake a strategic review of our business without a focused consumer research project. . . . We use consumer research in a variety of ways. First, to determine what consumer segments we should focus on. Second, to give us direct feedback on some of our latest concepts and merchandise programs. Third, to make sure that the consumers are the judge of what they want, as opposed to ourselves. We don't want to become overly expert and rely on our intuition instead of our customers'.[8]

---

[8] "Market Research Put to Creative Use," *Chain Store Age Executive*, May 1988, p. 24.

Marketing Concept
The philosophy of customer orientation urging firms to uncover customer needs first and then coordinate all their activities to satisfy those needs.

Conducting marketing research to understand customers and their requirements, as illustrated by the applications mentioned by Mr. Bluestein, is becoming widespread as an increasing number of firms embrace the ***marketing concept***—the philosophy of customer orientation urging firms to uncover customer needs first and then coordinate all their activities to satisfy those needs. The marketing concept, by emphasizing the need to gain a good understanding of customers, is also stressing the importance of marketing research in the early stages of formulating marketing strategies. Many firms have had major new-product failures because they neglected to research market opportunities and constraints before deciding to introduce the new products. A case in point is Levi Strauss & Company. Impressed by the popularity of the Levi name in the jeans market, the company introduced a line of high-priced, fashionable apparel under the same name, in competition with designer brand names. However, this strategy failed miserably. In one year alone the company's earnings dropped by 23%. In the words of Levi's own chief operating officer: "We've realized that just putting the Levi's name on something isn't enough to gain instant market acceptance." After this costly realization, the company decided to refocus attention on its basic jeans business.[9]

Developing and Implementing Marketing Plans

Marketing Mix
The set of product, promotion, price, and distribution (or place) decisions related to a firm's market offering.

A firm that wants to benefit fully from the opportunities it uncovers in the marketplace must develop an effective ***marketing mix;*** that is, it must make sound decisions about the nature of its product, the ways of promoting the product, the price it must charge its potential customers, and the means of making the product available to them. Good marketing research will identify whether a firm's marketing mix is effective enough to maximize the benefits (e.g., sales, profits, or market shares) to the firm from available opportunities. Some of the most successful new-product introductions have been preceded by extensive marketing research to aid in the development and implementation of one or more elements of their marketing mixes.

The introduction of Underalls by the Hanes Company (maker of the well-known L'eggs panty hose) was preceded by extensive marketing research to develop the product concept and promotional theme. Hanes tested several product concepts, using focus group research, and singled out the concept of panty hose with built-in underpants as the one with greatest appeal to customers. The company then developed prototypes of the built-in underpants garment and tested them in-house among its women employees. These in-house tests revealed several potential benefits that could be used to promote the product. Hanes further researched these benefits with an external sample of panty hose wearers and discovered that "appearance"—panty lines disappearing under slacks and dresses—was the most appealing benefit. Thus the highly successful promotional theme of "panties and panty hose all in one that would eliminate panty lines" emerged.[10] Another example of a product that benefited immensely from systematic marketing research is Ford Motor Company's hot-selling Taurus automobile. As the design for the car was being finalized, driving tests and interviews with hundreds of customers were conducted. This research has been credited with suggesting distinct and competitively superior features such as a net in the trunk to hold grocery bags upright, oil dipsticks painted a bright yellow for fast identification, comfortable seats, and easy-to-use dashboard instruments and controls.[11]

---

[9] "It's Back to Basics for Levi's," *Business Week,* March 8, 1982, p. 77.

Evaluating the Effectiveness of Marketing Plans

Controlling (Control Function) *Getting feedback from the marketplace and taking needed corrective action.*

Getting feedback from the marketplace and taking needed corrective action—often referred to as ***controlling,*** or the ***control function***—is an important component of the decision-making process. To be successful in the marketplace, a firm must at least periodically monitor market conditions (i.e., obtain feedback) and answer control-related questions like the following: What is the market share of our product? Is its share increasing, decreasing, or staying the same? Who are its users? Are the nature of the users and the volume of their purchases consistent with our expectations (goals)? If not, why not? A firm may want to explore a number of questions like these in order to evaluate its market performance. Accurate answers to such questions can only emerge from marketing research rather than from the intuition of marketers.

To illustrate the importance of marketing research in the control stage, consider the Mirro Company, a well-known producer of aluminum housewares. This company introduced a new line of microwave oven cookware under the Mirro name. The initial sales of this line were disappointingly low. Mirro's salespeople believed that the color of the cookware, which was a cocoa brown (the competitors' was beige), was to blame. However, instead of assuming that the salespeople's perceptions were accurate, Mirro conducted extensive consumer research to verify those perceptions. This research revealed that the Mirro name was the culprit—consumers associated *Mirro* with aluminum products that cannot be safely used in the microwave ovens. The research also showed that the cookware's color was *not* a drawback. Indeed, consumers were overwhelmingly *in favor* of the cocoa brown color because they perceived

---

[10] "Underalls' Success Due to 'Flanking Strategy,' Product Idea, Positioning," *Marketing News,* November 14, 1980, p. 11.

[11] "How Ford Hit the Bull's-Eye with Taurus," *Business Week,* June 30, 1986, pp. 69–70.

Exhibit 1.3 *AMA's Definition of Marketing Research*

Marketing research is the function which links the consumer, customer, and public to the marketer through information—information used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process.

Marketing research specifies the information required to address these issues; designs the method for collecting information; manages and implements the data collection process; analyzes the results; and communicates the findings and their implications.

Source: American Marketing Association, 250 South Wacker Drive, Chicago, Ill. 60606-5819. Used by permission.

it as being stain-resistant and having a quality image. Therefore Mirro changed the cookware's name to *Koolware* but left its color intact. The name change helped boost the cookware's sales.[12]

In summary, marketing research is an essential link between marketing decision makers and the markets they operate in. It can and should play an important role in all three stages of decision making: identifying marketing opportunities and constraints, developing and implementing marketing plans, and evaluating the effectiveness of marketing plans. The primary elements of marketing research and its potential contributions to decision making are captured in an expanded definition of marketing research developed by a task force of the American Marketing Association (see Exhibit 1.3).

PROVIDERS OF MARKETING RESEARCH SERVICES

In-House Marketing Research *Research conducted by a firm's own marketing research department or employee(s) in charge of marketing research.*

External Providers of Marketing Research *Research suppliers such as commercial marketing research firms and consultants.*

Firms can acquire the information relevant for marketing decision making through ***in-house marketing research*** (i.e., conducting the research by themselves) and/or through ***external providers of marketing research*** (i.e., contracting the research to outside agencies).[13]

In-House Marketing Research

In-house marketing research is either contained within a formal marketing research department or merely assigned as a responsibility of some employee(s). As one would intuitively expect, large firms are more

---

[12] "Mirro Changes the Recipe," *Sales & Marketing Management,* February 8, 1982, p. 13.

[13] See Richard D. Crisp, "Organization of the Marketing Research Function," in *Handbook of Marketing Research,* ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 1-63–1-72.

Research in Use 1.1 *Marriott Corporation*

Organization of Marketing Research

Marriott Corporation is a leading company in the services sector and employs well over 200,000 individuals. Although Marriott is perhaps best known for its lodging operations, it is also involved in a variety of food services operations consisting of food-contracting and restaurant services. In 1987 Marriott's lodging operations (e.g., Marriott Hotels and Resorts, Marriott Suites, Residence Inn by Marriott, Marriott Courtyard, and Fairfield Inns) accounted for about 41% of revenues. Food services operations (e.g., Host International, Inc., Marriott In-Flite Services, Roy Rogers restaurants, and Travel Plazas by Marriott) accounted for the remaining 59% of revenues. While Marriott's various businesses are linked by a common service component, each business serves a unique set of customers and therefore has a unique set of information needs.

At Marriott marketing research is primarily a corporate-level (i.e., centralized) function carried out by a department known as Corporate Marketing Services, or CMS. The CMS department consists of more than 30 research professionals. The department head, a vice president, reports directly to the senior vice president for strategic planning and business development who, in turn, reports to the chairman of Marriott.

Though Marriott's CMS is at the corporate level, it is organized to provide the benefits of a decentralized research function. Specifically, reporting directly to the VP of CMS are 10 directors, each responsible for meeting the unique research requirements of a particular business or set of businesses. To gain intimate knowledge of the unique needs of busi-

likely than small firms to have formal marketing research departments. Moreover, the incidence of such departments has been found to be higher in consumer product firms than in industrial-product firms and service firms.[14]

A key issue facing firms having an in-house marketing research department is how the department should be organized. In particular, should it be *centralized*—wherein one corporate department will cater to all the research needs of a firm—or should it be *decentralized*—wherein each "constituency" (e.g., division or group based on product line or market segment) will have its own marketing research arm? This issue is not easy to resolve because each mode of organization has distinct strengths that the other mode lacks.[15] For instance, a centralized or-

---

[14] Dik Warren Twedt, ed., *1983 Survey of Marketing Research* (Chicago: American Marketing Association, 1983).

[15] This is one of the key conclusions drawn by Crisp, "Organization of the Marketing Research Function."

Research in Use 1.1 *Continued*

nesses they are responsible for, each director has a "dotted-line" reporting relationship with senior managers of those businesses. Some directors also have one or more additional researchers reporting to them and working exclusively with the same businesses.

A key strength of this type of organization is that each major strategic business unit has a research director who knows the business and works on a regular basis with key clients in the business. Thus the director acts as a consultant to his or her business, proactively recommending needed research and suggesting actions warranted by the research findings.

Even though the directors have considerable autonomy, the CMS has several control and coordination mechanisms in place to achieve the benefits of centralized marketing research. For example, the directors must get prior approval from the VP of CMS for particularly controversial or high-impact research projects. Moreover, the VP of CMS reviews monthly status reports on all research conducted for the various strategic business units. To facilitate the achievement of cross-business synergies (e.g., sharing of knowledge, obtaining economies of scale), the 10 directors meet monthly with the VP. Also, each month, the VP holds meetings with the full research staff of the lodging and food services groups of businesses. These regular meetings encourage informal sharing of experiences and technical approaches, facilitate collaboration on difficult research problems, and reveal opportunities for operational efficiencies.

Source: This illustration is based on information contained in Frank E. Camacho and D. Matthew Knain, "Listening to Customers: The Market Research Function at Marriott Corporation," *Marketing Research: A Magazine of Management & Applications,* March 1989, pp. 5–14.

ganization will offer economies of scale. To have a corporate marketing research department with experts on questionnaire design, sample selection, statistical techniques, and so on, on its staff may be more cost-efficient than to duplicate such expertise at various decentralized areas. However, a centralized research function may not be as effective as a decentralized one in terms of understanding and responding to information requirements that are unique to each product line or market. To benefit from the strengths of both organizational modes, some firms, especially those involved in multiple lines of business, have adopted a *mixed* organizational structure wherein a corporate research function coexists with and complements a decentralized research function. Research in Use 1.1 describes the use of such a structure by Marriott Corporation.

External Providers of Marketing Research

Many organizations—those with and those without their own in-house marketing research capabilities—call upon external research agencies to meet at least some of their information needs. External providers of marketing research services consist primarily of commercial marketing research firms (including one-person consulting firms), academic consultants (i.e., marketing faculty members who offer consulting services on a part-time basis), and other agencies such as trade associations and the U.S. Census Bureau. Of the three types, commercial marketing research firms account for the majority of the total range of services offered by external providers. Hence we will discuss such firms in some detail.

According to Richard D. Crisp, commercial marketing research firms

> range from the very large, full-service marketing research agency qualified to handle a major assignment, from planning through delivery of final report and recommendations, to smaller firms that specialize in doing part of the total task. The latter category includes firms specializing in interviewing or other types of fieldwork, in conducting store audit studies as part of test market activities, or net product programs, etc.[16]

Approximately eleven hundred commercial marketing research firms operate in the United States.[17] These firms vary widely in terms of the geographic scope of their operations—ranging from international to local—as well as in terms of the range and nature of their services. To illustrate, the following firms are all external providers but offer different types of research services:

> **Burke Marketing Services, Inc.,** through its Burke International Research, Inc., and Burke Marketing Research divisions, offers a very wide range of custom-designed research services to client organizations all over the world.[18]
>
> **A. C. Nielsen Company** also operates internationally, but its offerings are dominated by syndicated services rather than custom-designed services (we will see more about syndicated services in Chapter 5 when we talk about secondary data).[19]
>
> **Namelab** is a limited-service firm specializing in developing and testing names for a variety of products and services.[20]
>
> **Legal Marketing Research, Inc.,** is a firm catering exclusively to legal clients. It offers "custom designed and legally oriented marketing research and consulting products for law firms and marketing entities

---

[16] Crisp, "Organization of the Marketing Research Function," pp. 1–66.

[17] Based on the listings in American Marketing Association, *International Directory of Marketing Research Houses and Services* (New York: American Marketing Association, 1986).

[18] American Marketing Association, *International Directory,* pp. 44, 46.

[19] American Marketing Association, *International Directory,* pp. 162–163.

[20] Further details about Namelab are available in a brochure entitled "NAMELAB—Name Development & Testing Laboratory," available from Namelab, 2229 Beach Street, San Francisco, Calif. 94123.

> where litigation may be involved. [Its services include] expert witness testimony, trademark and packaging studies, pricing and antitrust situations, advertising studies and others."[21]

> **Associates Interviewing Services** operates in a limited geographic area and primarily offers fieldwork services, as the following description indicates: "Philadelphia interviewing service covering entire metro area. High quality work. Personal, telephone, in-store interviewing, product placements, executive and in-depth interviews. Permanent suburban mall location with excellent demographics for group sessions, taste tests, package testing, advertising testing. One-way mirror. Kitchen."[22]

Though there are numerous marketing research firms in the United States, just a handful of them account for a large share of the total marketing research industry revenues. In other words, most of the research firms are relatively small. To illustrate, A. C. Nielsen Company, the largest research firm in 1988, had revenues of $880 million, which accounted for over 33% of the total revenues of the top 50 firms.[23] By contrast, the 10th largest firm had revenues of about $40 million and the 50th largest firm had revenues of only $5 million. The skewed size distribution within the marketing research industry results because most firms operate only in a small geographic area and offer only a limited range of research services.

Organizations without in-house marketing research capability invariably have to hire an external provider when they want information from their marketing environment. However, even firms capable of conducting research on their own can and do use external providers under certain circumstances (e.g., when specialized research services beyond the capabilities of in-house marketing research are needed). The appendix to this chapter contains a discussion of when to use external providers of research and how to select the most appropriate provider.

MARKETING MANAGER–MARKETING RESEARCHER INTERFACE

The usefulness of marketing research, whether conducted in-house or by external agencies, depends not only on how good the research is but also on how potential users perceive the research and apply its results. Marketing researchers and marketing managers (or users of marketing research) must work as a team in order to maximize the benefits derived from research. Does such teamwork exist in practice? Unfortunately, the answer to this question is not an unconditional yes. The relationship between marketing researchers and marketing managers is apparently not as effective as it should be.[24]

---

[21] American Marketing Association, *International Directory*, p: 119.

[22] American Marketing Association, *International Directory*, p. 30.

[23] "Top 50 U.S. Research Organizations," *Advertising Age*, June 5, 1989, p. S-4.

What factors are responsible for the less-than-ideal relationship between marketing researchers and marketing managers? An examination of the marketing literature on this issue reveals several general factors that contribute to a lack of effective teamwork between research providers and research users. These factors can be labeled as "complaints" that each party has against the other.

Let us first examine the complaints that managers, in general, have against researchers. A frequent complaint of managers is that *researchers are too technique-oriented;* that is, researchers are more interested in showing their sophisticated techniques than in ensuring and demonstrating the practical utility of their results. The following comments by a senior marketing executive succinctly reflect this concern:

> What we [managers] would like to see them [researchers] do is to define themselves as true staff support, *going beyond their technical specialties* [emphasis added], by using their accumulated knowledge and experience to help us. . . . Any technique or analytical approach we use must contribute to marketing decision making.[25]

A related complaint of managers is that *researchers do not communicate well* with them; researchers do not speak the manager's language. Owing to their obsession with techniques, researchers may tend to use overly technical terminology, thereby annoying managers and essentially shutting them out. Moreover, in their enthusiasm for describing their techniques and methodology, they may fail to adequately recognize and convey the *practical implications* of their findings. Here is what a senior vice president of marketing had to say regarding this problem:

> If professional researchers go beyond technique into diplomacy, persuasion, and statesmanship, the communication gap—the weakest link between researcher and management—will be bridged, and they will become a team that makes business succeed.[26]

Managers also believe that *researchers generate information of questionable value* and that *they fail to demonstrate the monetary benefits of their research.* Such a perception obviously jeopardizes managers' trust in marketing researchers and, consequently, widens the rift between them. For instance, a survey of 50 top corporate executives in the United States revealed that while a majority of them acknowledged the critical importance of marketing in strategic planning, only 16% gave an acceptable rating for the work done by marketing researchers.[27] This survey further showed that inefficiency and impractical results of marketing research were the main reasons for the corporate executives' poor opinions about it.

---

[24] For a discussion of the nature of and reasons for marketing managers' opinions about marketing research, see, for example, Danny Bellenger, "The Marketing Manager's View of Marketing Research," *Business Horizons,* June 1979, pp. 59–65; for an external supplier's views on corporate research users, see Joshua D. Libresco, "Do Bad Research, in 10 Quick and Easy Steps," *Marketing News,* January 4, 1988, pp. 4–5.

[25] "Technique-Obsessed Marketing Researchers Can't Help Managers Make Tough Decisions," *Marketing News,* January 22, 1982, Section 2, p. 1.

[26] "Successful Marketing Researchers Have Ability to Communicate, Breathe Meaning into Numbers," *Marketing News,* January 22, 1982, Section 2, p. 1.

Now let us look at researchers' complaints about managers. One major concern of researchers is that *managers do not provide enough funds for research.* According to one marketing research director:

> In many cases, research costs are held to absurdly low levels despite an extremely high value for the information to be obtained. Although the research conducted might be significantly less reliable or sensitive than a properly funded study, the data often are used in precisely the same manner.[28]

Indeed, several studies have shown that the funds allocated by management to marketing research are quite low in many firms.[29]

Low budget allocations for marketing research can adversely affect the marketing researcher–marketing manager interface in two ways:

1. Researchers may *perceive* the low allocations as demeaning their importance and self-esteem, which, in turn, may aggravate their already rough relationship with managers.
2. If low budget allocations lead to poor-quality research (as alleged by researchers), and if managers use such research results in their decision making, wrong decisions are likely to be made—decisions that managers can blame on the researchers, thereby creating further friction between the two parties.

A second general complaint of researchers is that *managers do not take the research results seriously;* that is, managers ignore marketing research results, or, worse still, use them only when they are consistent with their preconceived notions. To illustrate, Dale Pond, senior vice president of marketing at Payless Cashways, himself a research user, has observed: "Too often, I've seen market research used in much the same way the proverbial Saturday night drunk uses a lightpost—not for illumination but for support."[30]

---

[27] "Rabin: Top Execs Have Low Opinion of Marketing Research, Marketers' Role in Strategic Planning," *Marketing News,* October 16, 1981, p. 1.

[28] "Nine Research Mistakes Marketers Should Avoid," *Marketing News,* September 19, 1980, p. 14.

[29] See, for example, A. Parasuraman, "Research's Place in the Marketing Budget," *Business Horizons,* March/April 1983, pp. 25–29.

[30] Dale Pond, "My Opinion," *Home Center,* April 1987, p. 3.

This complaint is, perhaps, more factual than perceptual, because managers do request marketing research for reasons *other than* aiding their decision making. For instance, a manager may commission a research project simply to use it as a scapegoat in the event a decision (made *before* the research is done) turns out to be a bad decision. Such research has been labeled ***pseudo-research.***[31] Obviously, pseudo-research cannot be expected to foster a healthy, mutually trusting relationship between managers and researchers.

Pseudo-Research
Research requested by managers for reasons other than aiding their decision making.

Another complaint of researchers, which is not totally independent of the earlier complaints, is that *managers do not involve them in the decision-making process.* Researchers appear to feel that managers use them merely as data gatherers. Such a feeling, of course, helps neither the self-image of the researchers nor the researchers' opinion of the managers.

Having stated the general complaints of managers and researchers, we can make a couple of observations related to them. First, *both* researchers and managers may be responsible for the less-than-smooth relationship between them, which is a result of the perceptions of each group about the other group. For instance, managers may be reluctant to allocate funds for marketing research because they perceive researchers to be more technique-oriented than decision-oriented; however, researchers may *want* to be technique-oriented under the assumption that such an orientation may restore their self-esteem, which they perceive as being denied by managers. To the extent that such perceptual barriers can be removed through the joint and sincere efforts of managers and researchers, the relationship between them can be strengthened and made more productive.

Second, such adversarial relationships between researchers and managers certainly do not exist in *all* firms. Indeed, many firms today have overcome any interface problems and have successfully integrated marketing research with marketing management. For instance, advertising decision making in Sears is closely intertwined with marketing research.[32] Other companies whose effective integration of marketing research and management has been discussed in the literature include Marriott Corporation[33] and Amoco Oil.[34]

---

[31] For a discussion of this issue, see S. A. Smith, "Research and Pseudo-Research in Marketing," *Harvard Business Review,* 52 (March 1974), pp. 73–76.

[32] "Research Helps Sears Reap Optimum Impact from Ads," *Marketing News,* December 26, 1980, p. 7.

[33] Frank E. Camacho and D. Matthew Knain, "Listening to Customers: The Market Research Function at Marriott Corporation," *Marketing Research: A Magazine of Management & Applications,* March 1989, pp. 5–14.

[34] Abdul G. Azhari and Joseph M. Kamen, "Marketing Research at Amoco Oil: The Culture, the Principles, and the Contributions," *Marketing Research: A Magazine of Management & Applications,* June 1989, pp. 3–10.

ETHICS IN MARKETING RESEARCH

The purpose of this section is to expose you to certain key issues relating to the rights and obligations of all those involved in and affected by marketing research. Marketing research ethics seldom receive as much attention as other aspects of research. Nevertheless, ethical considerations pervade the practice of marketing research, often very subtly. In fact, the main reason for discussing research ethics early in this book is to encourage you to ponder the ethical questions implicit in many topics to be covered later.

tunkeutua/levitä

punnita

sisältyvät

Code of Ethics

Marketing Research Ethics *Rules and standards that govern the marketing research profession.*

The *Webster's II Dictionary* defines ethics as "the rules of standards of conduct governing the members of a profession."[35] From this definition ***marketing research ethics*** may simply be viewed as the *rules and standards that govern the marketing research profession.* In fact, several professional marketing organizations have their own written codes of marketing research ethics. For instance, Exhibit 1.4 contains the code of marketing research ethics drawn up by the American Marketing Association, whose members are professionals from academic and business organizations as well as a variety of nonbusiness organizations such as government agencies.

menettely/johto

ammattikunta

määritelmä

Codes of ethics can serve as useful general guidelines, and a genuine need for them exists in every profession. Empirical evidence suggests that the existence and enforcement of codes of ethics in research organizations are associated with higher levels of ethical behavior.[36] However, resolving ethical questions in specific research situations is not as simple as a written code may lead you to believe. The president of Creative Research Group Ltd. (Toronto), Richard Crosby, had this to say about codes of ethics:

aito

kokemusperäinen

voimaansaattaminen

> Various marketing research associations have codes of ethics, which implies that ethical matters have been nipped, cauterized even before the bud. The problem is that circumstances are so variable and individualized that they defy generalization. Debate over the need for a code of ethics is as old as the research profession and [my remarks] won't settle the matter. However, the case against a uniform ethical code is a strong one, based on the nature of the marketing research industry.[37]

tukahdetettu/poltettu alkuunsa

niitä on mahdoton yleistää

---

[35] *Webster's II New Riverside University Dictionary* (Boston: Houghton Mifflin, 1984).

[36] O. C. Ferrell and Steven J. Skinner, "Ethical Behavior and Bureaucratic Structure in Marketing Research Organizations," *Journal of Marketing Research,* 25 (February 1988), pp. 103–109.

[37] Richard Crosby, "Uniform Ethical Code Is Impractical Due to Shifting Marketing Research Circumstances," *Marketing News,* September 18, 1981, Section 2, p. 16.

Exhibit 1.4 *Marketing Research Code of Ethics of the American Marketing Association*

The American Marketing Association, in furtherance of its central objective of the advancement of science in marketing and in recognition of its obligation to the public, has established these principles of ethical practice of marketing research for the guidance of its members. In an increasingly complex society, marketing management is more and more dependent upon marketing information intelligently and systematically obtained. The consumer is the source of much of this information. Seeking the cooperation of the consumer in the development of information, marketing management must acknowledge its obligation to protect the public from misrepresentation and exploitation under the guise of research.

Similarly the research practitioner has an obligation to the discipline he practices and to those who provide support for his practice—an obligation to adhere to basic and commonly accepted standards of scientific investigation as they apply to the domain of marketing research.

It is the intent of this code to define ethical standards required of marketing research in satisfying these obligations.

Adherence to this code will assure the users of marketing research that the research was done in accordance with acceptable ethical practices. Those engaged in research will find in this code an affirmation of sound and honest basic principles which have developed over the years as the profession has grown. The field interviewers who are the point of contact between the profession and the consumer will also find guidance in fulfilling their vitally important role.

FOR RESEARCH USERS, PRACTITIONERS AND INTERVIEWERS

1. No individual or organization will undertake any activity which is directly or indirectly represented to be marketing research, but which has as its real purpose the attempted sale of merchandise or services to some or all of the respondents interviewed in the course of the research.
2. If a respondent has been led to believe, directly or indirectly, that he is participating in a marketing research survey and that his anonymity will be protected, his name shall not be made known to anyone outside the research organization or research department, or used for other than research purposes.

FOR RESEARCH PRACTITIONERS

1. There will be no intentional or deliberate misrepresentation of research methods or results. An adequate description of methods employed will be made available upon request to the sponsor of the research. Evidence that fieldwork has been completed according to specifications will, upon request be made available to buyers of research.
2. The identity of the survey sponsor and/or the ultimate client for whom a survey is being done will be held in confidence at all times, unless this identity is to be revealed as part of the research design. Research information shall be held in confidence by the research organization or department and not

used for personal gain or made available to any outside party unless the client specifically authorizes such release.

3. A research organization shall not undertake marketing studies for competitive clients when such studies would jeopardize the confidential nature of client-agency relationships.

FOR USERS OF MARKETING RESEARCH

1. A user of research shall not knowingly disseminate conclusions from a given research project or service that are inconsistent with or not warranted by the data.
2. To the extent that there is involved in a research project a unique design involving techniques, approaches or concepts not commonly available to research practitioners, the prospective user of research shall not solicit such a design from one practitioner and deliver it to another for execution without the approval of the design originator.

FOR FIELD INTERVIEWERS

1. Research assignments and materials received, as well as information obtained from respondents, shall be held in confidence by the interviewer and revealed to no one except the research organization conducting the marketing study.
2. No information gained through a marketing research activity shall be used directly or indirectly, for the personal gain or advantage of the interviewer.
3. Interviews shall be conducted in strict accordance with specifications and instructions received.
4. An interviewer shall not carry out two or more interviewing assignments simultaneously unless authorized by all contractors or employers concerned.

Members of the American Marketing Association will be expected to conduct themselves in accordance with the provisions of this Code in all of their marketing research activities.

Source: Marketing Research Standards Committee, *Marketing Research Code of Ethics* (Chicago: American Marketing Association). Used by permission.

yhtenäisenä

Furthermore, few, if any, ethical codes can be considered comprehensive in their coverage of topics concerning ethics in their respective professions.[38]

puutteellisuus/riittämättömyys

ahdinko

As another illustration of the inadequacy of codes of ethics, consider the plight of the Ethics Committee of the American Marketing Association's New York Chapter, which attempted to choose a code of ethics that would best address the chapter members' requests for ethical guidance on research issues: "Committee members produced 15 codes, none of which was suitable. The standards were either too broad and vague, too focused on a particular function of research, or too limited to the interest of a given organization.[39]

---

[38] For criticisms of codes of ethics within the marketing profession, see, for example, Alice M. Tybout and Gerald Zaltman, "Ethics in Marketing Research: Their Practical Relevance," *Journal of Marketing Research,* 11 (November 1974), pp. 357–368; and Shelby D. Hunt, Lawrence B. Chonko, and James B. Wilcox, "Ethical Problems of Marketing Researchers," *Journal of Marketing Research,* 21 (August 1984), pp. 309–324.

epämääräinen

An additional limitation of a code of ethics is that any so-called standard in an ethical code is in reality subject to different interpretations stemming from the values and priorities of the individuals seeking guidance from the code. Research ethics "are guidelines and principles that help us uphold *our values*—to decide which goals of research are most important and to reconcile values and goals that are in conflict"[40] (emphasis added).

rajoitus
riippuvainen
tulkinnat
juontavat alkunsa
sovittaa

C. Merle Crawford conducted an interesting study that found evidence of how individuals' value systems and interpretations can influence their perceptions about what is ethical and what is not.[41] Crawford presented, to a sample of marketing managers and researchers, several hypothetical research scenarios, each involving an action that was potentially questionable on ethical grounds. The survey respondents were asked to state whether they approved or disapproved of the action taken in each scenario and to give reasons for their answers. One of the scenarios was the following:

käsityksiin

> One product of the X Company is brassieres, and the firm has recently been having difficulty making some decisions on a new line. Information was critically needed concerning the manner in which women put on their brassieres. So the Marketing Research Director designed a study in which two local stores cooperated in putting one-way mirrors in their foundations dressing rooms. Observers behind these mirrors successfully gathered the necessary information.[42]

About 80% of the managers and researchers disapproved of the use of one-way mirrors in this situation. As one might expect, the predominant reason for their disapproval was their perception of the action as a serious invasion of privacy. However, 20% of the respondents *approved* the action. Apparently they assumed that the observers were female and hence they felt that the need for information was much more critical than the issue of moral infractions. One of these respondents is reported to have said, "The women don't know they've been observed, and thus can suffer no mental anguish." How do you feel about the appropriateness of using one-way mirrors in this scenario?

therefore
rikkomukset
henkistä ahdistusta

---

[39] Marilyn Landis (Hauser), "Why It Took Three Years for Ethics Panel to Polish Pledge," *Marketing News*, September 19, 1980, p. 1.

[40] Edward Diener and Rick Crandall, *Ethics in Social and Behavioral Research* (Chicago: University of Chicago Press, 1978), p. 3.

[41] C. Merle Crawford, "Attitudes of Marketing Executives Toward Ethics in Marketing Research," *Journal of Marketing*, 34 (April 1970), pp. 46–52.

[42] Crawford, "Attitudes of Marketing Executives," p. 47.

In sum, perceptions about what is ethical clearly depend on one's value system and subjective judgment.

Ethical Issues Concerning Various Parties

Several different parties are directly or indirectly associated with marketing research. We can classify them into five groups:

ulkopuolinen

1. External research providers (typically full-service firms) that deal directly with client organizations; these firms are traditionally known as *research suppliers.*
2. Limited-service external providers (e.g., data-gathering firms) that cater to research suppliers as well as in-house marketing research departments; these firms are typically referred to as *field service firms.*
   [illegible]
3. *Research subjects* or respondents who provide the data collected.
4. *Clients* who are the users of marketing research information.
5. The *general public* that may be affected by the dissemination of research results.
   levittäminen

Each of the five groups has certain *rights,* which implies that each also has certain *obligations* toward the others. Examples of the rights of each party are discussed next.

Research Suppliers. Research suppliers have a right to be treated fairly by clients and by field service firms, the two groups they most frequently deal with. Alternatively, clients and field service firms have an obligation to treat research suppliers fairly. What does *fair treatment* mean? It can refer to a variety of ethical issues depending on the nature of the situation and values of the individuals involved. Let us look at an example.

> **Example.** Client, Inc., hires Research, Inc., to conduct a study to determine the extent of use of Client's brand (say, brand X) versus the use of competing brands. Research, Inc., conducts a study and presents the following conclusion to Client, Inc.: "Of the 2000 respondents surveyed 1000 refused to disclose the brand used; of the remaining 1000 respondents 500 used brand X, 300 used brand Y, and 200 used a variety of other brands." From this conclusion Client, Inc., develops the following ad campaign: "One out of every two consumers uses brand X, according to a study conducted by Research, Inc."
>
> paljastaa

What are the ethical issues involved in this example? Client has conveniently ignored the half of the sample that refused to disclose the brand used. Hence the ad claim is certainly questionable, unless Client has additional evidence suggesting that respondents to the study did not significantly differ from nonrespondents or that the 1000 respondents were similar to target market consumers. In the absence of such evidence the ad may also tarnish Research's reputation since it explicitly identifies

sopivasti

jättänyt ottamatta huomioon

tahrata

nimenomaisesti

Research as the source of the claim. Other knowledgeable industry sources, such as Client's major competitors, which may be aware that Client's ad claim is not true, are likely to think poorly of Research's methodology and hence its image. And what about the rights of the general public that will be exposed to the ad campaign? We will discuss more on this issue later.

Although the situation described depicted in the example is fictitious, similar happenings are apparently not uncommon in real life. For instance, according to Nancy Finn, group marketing manager for Clairol, many marketers deliberately vääristävät distort research findings. Finn labels them "data manipulators," who "have little respect for the panos contribution of an individual researcher or the oppiaine discipline of marketing research."[43]

Field service firms also have certain obligations to research suppliers. For instance, a field service firm that poikkeaa deviates from the data collection procedures laid down by a research supplier, *without the suostumusta consent of the latter*, is not being ethical. Yet such violations apparently occur frequently. As an executive of Data Development Corporation (a research supplier in New York), calling for greater professionalism, honesty, and accuracy on the part of field services, says: "We don't get the quality product we would like. . . . We trust field supervisors, yet it is in this area that my expectations suffer the most."[44]

Field Service Firms. One of the rights of field service firms is to receive adequate respect and compensation for their fieldwork. This statement almost sounds like a selviö truism. However, in practice, adequate compensation is not always received.[45] Consider the following scenario, which is claimed as being typical of the experiences of many field service firms:

> A field service firm on the West Coast receives a rush order for a long-distance telephone study from a client firm on the East Coast. Since time is short, a *verbal contract* is agreed to over the telephone, with the understanding that the field service will be korvataan reimbursed for all the study expenses, as long as the study is completed on time. The field service firm completes the job on time and mails the findings to the client along with a bill for the study's costs. However, the client questions certain charges on the bill, such as certain long-distance telephone charges and the charge for a field supervisor's time, and refuses to pay the bill in full. Instead, the client only pays what it feels is fair. The field service firm feels cheated and is frustrated at being unable to collect from an out-of-state firm.[46]

---

[43] Quoted in "Marketers Are Information Lovers, Expediters, or Manipulators When Interfacing with Researchers," *Marketing News*, January 22, 1982, Section 2, p. 12.

[44] Quoted in "Marketing Research Clients, Suppliers, Field Services Tell Expectations of One Another," *Marketing News*, September 18, 1981, Section 2, p. 3.

[45] "Marketing Research Clients, Suppliers, Field Services"; see also "Marketing Research Suppliers, Clients Should Respect the Professionalism of Data Collectors," *Marketing News*, January 22, 1982, Section 2, p. 2.

[46] Adapted from "How Unethical Firms 'Prey' on Marketing Research Suppliers," *Marketing News*, September 18, 1981, Section 2, p. 16.

Is the client firm being ethical? The field service firm would definitely not think so. Since there was no written contract, however, the ethical question is very hard to resolve objectively. Possibly, the conflicting viewpoints of the two parties arose owing to a genuine misunderstanding about what was covered by the verbal contract. This scenario is illustrative of the many gray areas in marketing research ethics; only the expectations, value systems, and sense of fairness of the particular parties involved can adequately resolve such issues.[47]

Research Subjects. Many modern-day forces such as the consumer movement and the increasing importance of individualism are placing the rights of research subjects in the limelight. Of particular interest to marketing research practice are the following rights of research subjects: the *right to privacy,* the *right to be informed,* the *right to decline to participate* in a study, the *right to physical and psychological safety,* and the *right to confidentiality* of responses. Although the American Marketing Association's code of ethics (Exhibit 1.4) does not explicitly address these rights, the codes of other associations, like the American Psychological Association, do so.[48] These rights are self-explanatory, and any concerned researcher should be aware of them and be willing to respect them. However, looking beyond the mere definitions of these rights, one will see numerous ethical gray areas in which a researcher will have to make judgment calls.

parrasvaloihin

itsestäänselvä

pelkät määritelmät

Take a survey respondent's right to privacy, for instance. All of us will probably agree that telephoning a respondent in the middle of the night as part of a survey is a serious violation of the respondent's right to privacy. What about calling the respondent about 7 p.m.? You will probably say, "That's okay," But see if you still think so after reading the following example.

Example. It is 7 p.m. and John Hagen is right in the middle of what he considers to be a very important dinner with others. The phone rings, and the caller is conducting a survey. Says the caller: "Mr. Hagen, you are one of a scientifically chosen few to participate in this *extremely important* study dealing with. . . . It is critical that you take just a few minutes *now* to respond to my survey. Your refusal to participate now will seriously affect the usefulness of the study. Won't you *please* help us by participating?"

Should John agree to participate? If he does, his important dinner is going to be further interrupted; moreover, depending on the nature of the survey questions, his mood may be affected. The result can be viewed as an unfair invasion of John's privacy. What if John simply exercises his right to decline to participate and hangs up? This action will prevent any further violation of his privacy; but will it also make him feel guilty and cause him psychological pain? If it does, his right to safety will be violated to a certain degree.

---

[47] Other illustrations of ethical conflicts in client-researcher dealings can be found in Sidney-Hollander, Jr., "Ethics in Marketing Research," in *Handbook of Marketing Research,* ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 1-107–1-127.

[48] *Ethical Principles in the Conduct of Research with Human Participants* (Washington, D.C.: American Psychological Association, 1973).

Does the researcher have any ethical responsibility here? Should, or could, the researcher have anticipated a situation like this and taken precautions to prevent it from happening?

Although the situation in this example is hypothetical, and perhaps even somewhat extreme, it is illustrative of real ethical issues researchers wrestle with frequently. A *broad* ethical question underlying scenarios such as the one presented is, Where does one draw the line between a researcher's need (or right) for information and the rights of research subjects? There are no easy solutions to such questions.[49]

Clients. Client organizations have a right to expect fair treatment from the other parties in the marketing research arena. For instance, they have a right to get an accurate, honest estimate of the cost of a study to be done by a research supplier. This right appears to be straightforward. However, consider this situation: A research supplier estimates that a study will cost $10,000, and the client agrees that this price is reasonable. But when the study is completed, the research supplier submits a bill for $15,000, claiming that the cost increased because of changes the client wanted made during the course of the study. The client, while acknowledging that certain unplanned changes were made, argues that such changes should have cost no more than an additional $2000. Who is to be blamed here? This question, too, is an ethical issue with no easy answer.

Clients also have a right to expect research suppliers to be honest and accurate in meeting the clients' information needs. Research practice indicates again that this right is not easily achieved. The nature of the research industry is such that research suppliers may be tempted, or even forced, to compromise their integrity in their quest for business survival. For instance, here is what Harry Heller, president of Harry Heller Research Corporation, a research supplier, had to say about his industry: "The proliferation of research suppliers has led to the basic, underlying problem in our field. Suppliers are over-promising. As a result, clients are over-demanding. . . . There are more than enough companies that will promise anything.[50]

---

[49] For a related discussion, see Leo Bogart, "The Researcher's Dilemma," *Journal of Marketing,* 26 (January 1962), pp. 6–11.

[50] Quoted in "Honesty, Better Quality Needed from Research Suppliers, Field Services," *Marketing News,* September 18, 1981, Section 2, p. 4.

General Public. The rights of the general public are not very different from those of research subjects. The only major difference between the two sets of rights is that while the rights of the research subjects become relevant *during data collection,* the rights of the general public become relevant *if and when any research information is disseminated.*

Recall the scenario involving Client, Inc., and Research, Inc., that we looked at while discussing the rights to research suppliers. In that scenario Client made a questionable ad claim, which we said was unfair to the reputation of Research. By using the claim in a national ad campaign, Client would also violate the right of the general public to be informed accurately. This action is a clear case of unethical behavior on the part of Client. Unfortunately, ethical matters concerning the rights of the general public are not always as clear-cut, as the next example indicates.

> **Example.** During the 1980 presidential election on November 4, NBC, ABC, and CBS interviewed numerous voters in order to predict the winner. On the basis of its interviews NBC was the first to predict that Ronald Reagan would be the winner. NBC made its prediction public before its competitors and *before the polls closed on the West Coast.*[51] NBC certainly had the right to inform the public and the right to stay ahead of its competition. But were any of the general public's rights violated? Specifically, did NBC violate the rights of any West Coast voters by announcing a predicted winner *before* they had a chance to vote?

The area of ethics in marketing research is extremely important and concerns the rights of a variety of parties. Nevertheless, the resolution of real-life ethical conflicts is not always easy. Potential ethical conflicts can perhaps be minimized, but not necessarily eliminated, if the three parties in the research industry—research suppliers, field service firms, and client organizations—make a conscious effort to be aware of and respect the rights of one another as well as those of the research subjects and the general public. This section presented a number of situations involving ethical questions but, by necessity, left many of those questions unanswered. However, the objective of this section can be considered to have been accomplished if it has sensitized you to be concerned about marketing research ethics.

SUMMARY

The purpose of this chapter was to provide an overview of the nature and scope of marketing research. Marketing research is more than techniques for gathering data. Its purpose is to generate *information* that can aid decision makers in a variety of business as well as nonbusiness settings. Marketing research can provide useful inputs at all stages of the decision-making process: identifying marketing opportunities and constraints, developing and implementing marketing plans, and evaluating the effectiveness of those plans.

---

[51] "Peacock's Night to Crow," *Newsweek,* November 17, 1980, p. 82.

Marketing research can be conducted *in-house* or through an *external provider*. In-house marketing research ranges from a part-time responsibility of an employee to a separate department. External providers consist of commercial marketing research firms, academic consultants, and other associations or agencies. Of these, commercial marketing research firms are most frequently used by client organizations.

The benefits realized from marketing research depend on the degree of teamwork between research providers and research users. Unfortunately, there are apparently some barriers to this teamwork, although such barriers (especially perceptual barriers) can be, and have been, successfully overcome through mutually trusting dialogue and cooperation between the two parties.

Ethical questions pervade the field of marketing research and concern five different parties: *research suppliers, field service firms, research respondents, clients,* and the *general public.* A number of marketing organizations have codes of ethics. Unfortunately, a code of research ethics is merely a set of *general* guidelines. What is considered ethical in any given situation depends on the value systems and subjective judgment of the parties involved.

KEY TERMS

Marketing research
Information
Marketing concept
Marketing mix
Controlling (control function)
In-house marketing research
External providers of marketing research
Pseudo-research
Marketing research ethics

QUESTIONS

1. Critically evaluate the soundness of the following definition of marketing research: "Marketing research is a set of data collection techniques that can be used by marketers to help them in their decision making."
2. What lessons about marketing research are evident in the Edsel automobile example?
3. Name a *specific* purpose for which each of the following may be able to use marketing research fruitfully.
   a. A hospital,
   b. A United States congressman,
   c. A gas station,
   d. The Internal Revenue Service.

4. What are the three major components of marketing decision making to which marketing research can contribute?
5. The chief executive officer (CEO) of a firm tells top management: "Our sales have been increasing at a snail's pace over the past five years. Therefore I want our sales to increase by *at least* 15 percent during the next year. Do whatever is necessary to accomplish this." Develop a logical argument that can be used to show the CEO why the stated sales goal may be unrealistic.
6. Briefly discuss a specific example to illustrate the importance of marketing research in determining the corrective action to be taken when results from the marketplace are disappointing.
7. State the two alternative approaches that a firm can use to acquire marketing research information.
8. Would you consider each of the following a full-service research firm or a limited-service research firm? For each firm, indicate your reason in a sentence or two.
   a. Burke Marketing Services, Inc.,
   b. Legal Market Research, Inc.,
   c. Associates Interviewing Services.
9. "When a firm obtains bids from several external research suppliers for the *same* research project, it should hire the lowest bidder." Discuss this statement.
10. State the complaints that managers typically have against marketing researchers.
11. State the complaints that marketing researchers typically have against managers.
12. "The rift between researchers and managers is too wide to be bridged effectively." Discuss this statement.
13. Briefly discuss why a written code of ethics is not a panacea for resolving ethical questions in marketing research.
14. List the various parties that can be potentially involved in or affected by ethical issues in marketing research.
15. Are there any ethical issues implicit in the Edsel automobile example we discussed? If so, briefly discuss them.
16. Which two parties are most closely and frequently associated with research suppliers? Give examples to illustrate the obligations of these two parties toward research suppliers.

4. What are the three major components of marketing decision making to which marketing research can contribute?
5. The chief executive officer (CEO) of a firm tells top management: "Our sales have been increasing at a snail's pace over the past five years. Therefore, I want our sales to increase by at least 15 percent during the next year. Do whatever is necessary to accomplish this." Develop a logical argument that can be used to show the CEO why the stated sales goal may be unrealistic.
6. Briefly discuss a specific example to illustrate the importance of marketing research in determining the corrective action to be taken when results from the marketplace are disappointing.
7. State the two alternative approaches that a firm can use to acquire marketing research information.
8. Would you consider each of the following a full-service research firm or a limited-service research firm? For each firm, indicate your reason in a sentence or two.
    a. Burke Marketing Services, Inc.
    b. Leger Market Research, Inc.
    c. Associates Interviewing Service
9. "When a firm obtains bids from several external research suppliers for the same research project, it should hire the lowest bidder." Discuss this statement.
10. State the complaints that managers typically have against marketing researchers.
11. State the complaints that marketing researchers typically have against managers.
12. "The rift between researchers and managers is too wide to be bridged effectively." Discuss this statement.
13. Briefly discuss why a written code of ethics is not a panacea for resolving ethical questions in marketing research.
14. List the various parties that can be potentially involved in or affected by ethical issues in marketing research.
15. Are there any ethical issues implicit in the Edsel automobile example we discussed? If so, briefly discuss them.
16. Which [illegible] frequently associated with research suppliers? Cite examples to illustrate the obligations of [illegible] toward research suppliers.

Guidelines for Using External Research Providers

As mentioned in this chapter, even firms with in-house research capability sometimes employ the services of external research providers. When would such a firm be likely to benefit by engaging the services of an external research provider? Several citations in the marketing literature contain detailed lists of circumstances in which a firm should contract for outside research services.[1] Although some minor discrepancies appear across all these lists, they all suggest the same *general* factors to be considered. As shown in Exhibit A.1.1, these factors can be labeled as credibility, competence, cost, and capacity. Each factor is discussed next.

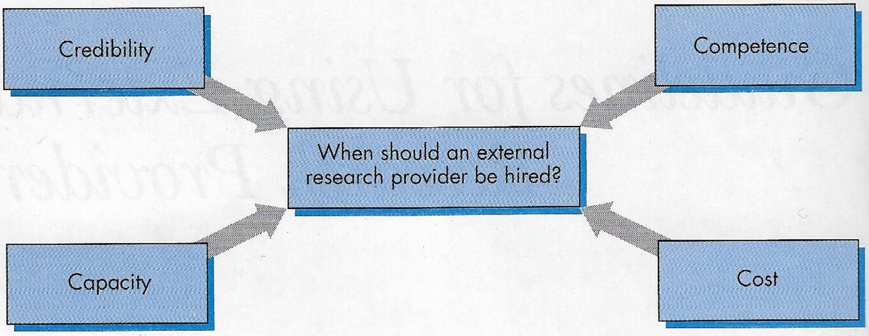
CREDIBILITY

Credibility refers to the perceived trustworthiness of the research and its findings. Even firms with full-fledged in-house marketing research departments hire outside agencies to conduct research in situations where the credibility, and hence usefulness, of the research may depend as much on *who* does the research as on *how well* it is done. For instance, suppose the board of directors of a large corporation wants to know how successful top management has been in building a favorable image for the corporation in the eyes of the public. In this situation the board is likely to find an assessment of corporate image made by an independent research firm more credible than an assessment based on data gathered by in-house marketing research. In general, the use of external providers is desirable when there is a possibility for conflict of interest (real or perceived) if a

---

[1] For instance, see R. A. Petersen and R. A. Kerin, "Effective Use of Marketing Research Consultants," *Industrial Marketing Management,* 9 (February 1980), pp. 69–73; R. Paul, "When Does Your Company Need Help from Marketing Research Consultants?" *Industrial Marketing,* 63 (August 1978), p. 48; L. Adler, "How to Tell If You Need the Services of an Outside Research House," *Sales & Marketing Management,* 20 (April 3, 1978), p. 74.

Exhibit A.1.1 *Factors Relevant in Deciding Whether to Use External Providers*



research study is conducted in-house—that is, when independence of the source of the research study is crucial in ensuring that the study's findings are credible to their intended audience.

COMPETENCE

Competence refers to special capabilities or facilities of an external research provider that an in-house marketing research department may not possess. A firm will have to hire an appropriate external agency for projects requiring special capabilities. For instance, consider A. C. Nielsen Company's National ScanTrack Service. This service is based on a nationwide sample of high-volume, scanner-equipped supermarkets, and it provides weekly data on product class, brand and item sales, brand share, selling prices, and so forth, for numerous packaged goods.[2] Nielsen is able to obtain such data readily because of its *wide experience* in store auditing and its *established contacts* with supermarkets. An individual packaged-goods manufacturer, even one with a good in-house marketing research department, will have difficulty gaining access to the kinds of data Nielsen is capable of obtaining.

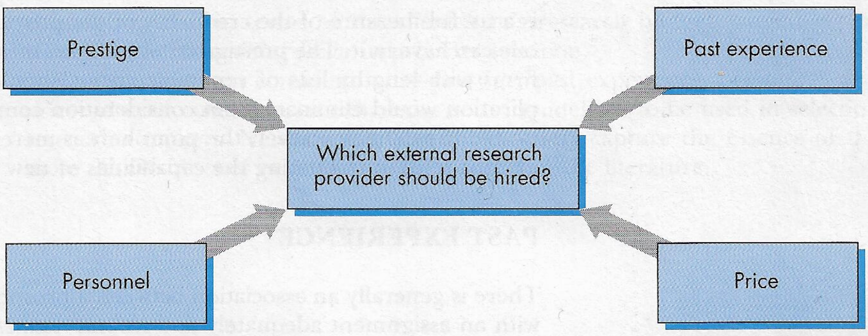
COST

With certain types of research projects, for a firm to use an external research provider to do part or all of the work may be more economical than for the firm to do the entire project on its own. External providers are especially economical when they specialize in the type of research work to be done. For instance, many in-house marketing research departments may *design* research projects but may hire a field service firm (such as the Associates Interviewing Services mentioned earlier) to perform the *data collection.* Maintaining a full staff of qualified field-

---

[2] Nielsen National ScanTrack Service—For Continuous Planning and Evaluation of Promotions," brochure prepared (in 1982) by A. C. Nielsen Company, Nielsen Plaza, Northbrook, Ill. 60062.

Exhibit A.1.2 *Criteria for Evaluating External Providers*



work professionals such as interviewers and data coders may be prohibitively expensive for an in-house research department.

CAPACITY

Sometimes, a firm may hire an external provider to conduct an important research project—one that cannot be delayed—when the firm's in-house research department is fully tied up with ongoing research projects. In other words, external providers may be employed to expand the *capacity* of in-house marketing research temporarily to meet urgent research needs. When the timing of a research project is crucial, a firm may be forced to secure the services of an external provider, even though in-house marketing research might be capable of handling the project *if* it had the time. The critical issues is, "*How soon* can in-house marketing research conduct the project?" rather than, "Is in-house marketing research *qualified* to conduct the project?"

When a firm decides to use external research services, it also has to decide which specific provider to hire. What criteria should a firm consider in selecting an external research provider? Several authorities have offered lists of criteria.[3] Fortunately, the various lists agree to a great extent and lead to a handful of general guidelines. These guidelines are reflected in the four selection criteria summarized in Exhibit A.1.2: prestige, past experience, personnel, and price. Each criterion is discussed next.

PRESTIGE

Prestige refers to the reputation of a research provider and is a surrogate indicator of its research quality and capabilities. Like any other profession, the research industry does have charlatans. Hence a firm should check out the reputation of prospective research suppliers. A simple way is to ask the suppliers for references or for a list of clients they serve. The reputation of a provider's clients is a useful measure of the credibility of the provider's work and the confidence one can have in it. The prestige criterion is *not* meant to imply that only research firms with lengthy lists of reputable clients should be considered; such an implication would eliminate from consideration competent *new* firms with sophisticated capabilities. Rather, the point here is merely that a client firm should be extra careful in evaluating the capabilities of new research firms.

---

[3] For instance, see C. S. Mayer, "Evaluating the Quality of Marketing Research Contractors," *Journal of Marketing Research*, 4 (May 1967), pp. 141–143; J. H. Myers, "Competitive Bidding for Marketing Research Services?" *Journal of Marketing*, 33 (July 1969), pp. 40–45; L. Adler, "How Marketing Research Helps Sales," *Sales & Marketing Management*, 115 (November 3, 1975), pp. 105–107.

PAST EXPERIENCE

There is generally an association between a prospective provider's ability to deal with an assignment adequately and its *past research experience,* especially experience related to the research assignment under consideration. Hence past experience is an important indicator, though not necessarily the *only* indicator, of a provider's competence. Another facet of past experience is the extent and nature of the working relationship the client firm may have had with the provider. Rehiring the services of a provider with whom past experiences have been good may offer certain advantages, such as lower cost and better client-provider understanding.

PERSONNEL

A key consideration in choosing an external research firm is the set of skills possessed by the firm's *personnel,* particularly those employees who will be in direct contact with client personnel. Here the term *skills* is not restricted to *technical* qualifications; quite to the contrary, it includes *nontechnical* skills such as interpersonal skills, communication skills, the ability to understand a client's research problem, and the ability to establish rapport with client personnel. The potential benefits of research conducted by an external provider will be diminished if there is not enough communication between client and provider personnel. The importance of maintaining an effective client-provider interface is underscored by the numerous books that have been written on this subject.[4]

PRICE

The *price* of a research firm's services is of obvious concern to any client firm. However, the cheapest buy is not necessarily the best buy. That is, a research firm with a good reputation and high-quality research capabilities usually can, and most likely will, demand a premium price. Also, competing providers may bid different prices for the same research assignment because their perceptions of the nature and scope of the assignment may be different.[5] In fact, a *written*

---

[4] See, for example, Chip R. Bell and Leonard Nadler, eds., *The Client-Consultant Handbook* (Houston: Gulf, 1979); John J. McGronagle, Jr., *Managing the Consultant—A Corporate Guide* (Radnor, Pa.: Chilton, 1981).

[5] For a detailed discussion of how marketing research suppliers set prices, see J. B. Haynes and P. L. Wilkens, "Pricing of Marketing Research Services," *Business Horizons,* 17 (October 1974), pp. 75–80.

proposal, detailing what will and will not be covered in a research assignment, should be developed and agreed to by both the client and provider *before* a final research contract is signed.[6] In other words, the price quoted by a research firm, although critical in terms of the client firm's research budget, should *not* be viewed as the *determining factor* in provider selection.[7]

The four general considerations—prestige, past experience, personnel, and price—do not constitute an exhaustive set of guidelines to be used in selecting research suppliers. Nevertheless, they do concisely capture the essence of the more detailed lists of evaluation criteria found in the literature.

---

[6] See A. B. Blankenship and R. F. Barker, "The Buyer's Side of Marketing Research," *Business Horizons,* 16 (August 1973), pp. 73–80.

[7] See J. B. Haynes and J. T. Rothe, "Competitive Bidding for Marketing Research Services: Fact or Fiction?" *Journal of Marketing,* 38 (July 1974), pp. 69–71.

CHAPTER 2

Marketing Research & Decision Support Systems

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between an ad hoc research project and systems-oriented marketing research.*
- *Define marketing information system, and describe its basic components.*
- *Define marketing decision support system, and distinguish it from a marketing information system.*
- *Describe what a fourth-generation language is.*
- *Illustrate the role of marketing research in developing marketing decision support systems.*
- *Define and provide an example of single-source data base.*
- *Define expert system, and discuss its role in marketing decision making.*

In Chapter 1 we saw that marketing research can provide valuable inputs at three basic stages of the marketing decision-making process: identifying marketing opportunities and constraints, developing and implementing marketing plans, and evaluating the effectiveness of marketing plans. The following are examples of questions that marketing research can help answer at each stage.

toteuttaa

- *Identifying marketing opportunities and constraints.* What is the market potential for our product? What are the demographic and socioeconomic characteristics of the region where we wish to market our product? How many potential competitors are there in the region and how intense is the competition among them?

ankaraa/kovaa

- *Developing and implementing marketing plans.* In what package sizes should we offer our product? What media are most appropriate for promoting our product? Through what types of retail outlets should we sell our product?
- *Evaluating the effectiveness of marketing plans.* How effective is our current promotional strategy? To what extent are users of our product inclined to buy it again? What are the market shares of our product and the products of our leading competitors?

Ad Hoc Research Project
A discrete, situation-specific project that is initiated and completed in response to a particular question, or set of related questions, raised by a decision maker.

The traditional role of marketing research has been to answer such questions through *ad hoc* research projects. An ***ad hoc research project*** is a discrete, situation-specific project that is initiated and completed in response to a particular question, or set of related questions, raised by a decision maker. Much marketing research is still performed in the form of discrete projects. However, during the past decade the role of marketing research in decision making has been gradually expanding. Mar-

erillinen

Systems-Oriented Marketing Research *Making available appropriate marketing information on a regular, integrated basis.*

keting research is increasingly shifting from being purely project-oriented to being systems-oriented. ***Systems-oriented marketing research*** involves making available appropriate marketing information on a regular, integrated basis. The shift toward a systems orientation is being fueled by two related forces: (1) a growing need for marketers to constantly monitor market developments, which have been occurring more and more rapidly; and (2) an opportunity to make use of advances in information technology to gain a competitive advantage.[1] John Webber, director of marketing research for General Foods Corporation, describes the nature and implications of the movement toward systems-oriented marketing research as follows:

> [I see a shift] from *ad hoc* information gathering to the development of marketing information and research *systems*. . . . We are increasingly seeing value in using a systems approach. By a systems approach I mean making sure that each piece of information, each piece of research investigation integrates and builds upon other information into relationships so that it can be used in a coherent overall framework. This is not to say that there will be no more ad hoc research studies. It is to say that those *ad hoc* studies will need to build into a coherent framework of knowledge about the business and its relationship to the marketplace. Accessing that knowledge base via information technology will reduce the need for special studies.[2]

John Webber's predictions imply that in coming years managers will continue to decrease their reliance on *ad hoc* research projects to answer their market-related questions. Instead, they will have increasing access to "frameworks of knowledge" from which they can readily extract the information they need to assist them in their decision making. The role of marketing research will expand to include assisting in the development of such frameworks and serving as a constant link between them and the marketplace. Therefore marketing research will continue to be a critical component in the construction and updating of the frameworks. The purpose of this chapter is to provide an overview of different types of knowledge frameworks or systems for aiding managers.

EVOLUTION OF SYSTEMS FOR SUPPORTING DECISION MAKING

The foregoing discussion pointed out that the manner in which managers are using marketing research is undergoing a transformation from a project orientation to a systems orientation. This transformation

---

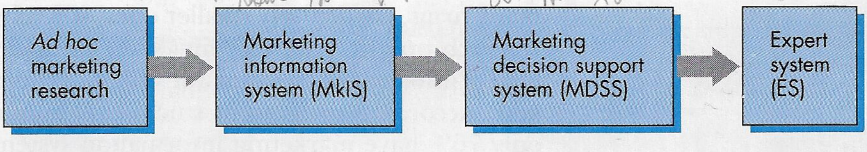
[1] These two forces are derived from a quote attributed to Lou Wallis, senior research associate for the Conference Board, in Thayer C. Taylor, "Sales Support Systems in the Computer Age," *Sales & Marketing Management,* March 1989, pp. 72–73.

[2] Excerpted from a speech given by John C. Webber, "Nine Shifts Changing the Nature of Marketing Research," Advertising Research Foundation 33rd Annual Conference, New York Hilton Hotel, March 2–4, 1987.

Exhibit 2.1 *Evolution of Systems for Supporting Decision Making*

regular/routine source of inform.

allows managers to interact


Ad hoc marketing research
Marketing information system (MkIS)
Marketing decision support system (MDSS)
Expert system (ES)

process is still evolving and resulting in increasingly sophisticated systems. Exhibit 2.1 summarizes the evolution of systems for supporting decision making.

Formal definitions of marketing information system (MkIS), marketing decision support system (MDSS), and expert system (ES) are given in subsequent sections devoted to each type of system. This section offers a preliminary glimpse of what each system is and how it differs from the preceding system. An MkIS is a system that provides managers with relevant marketing inputs on a regular, systematic basis. A key distinction between an MkIS and *ad hoc* marketing research is that the former is a *regular* or *routine* source of market information. An MkIS keeps managers in constant touch with the market by automatically channeling appropriate reports to them on a timely basis. An *ad hoc* research project, in contrast, gives managers a snapshot of a particular facet of a market in response to a specific request for information.

An MDSS is basically an MkIS that permits managers to request special types of data analyses or reports on an as-needed basis. In contrast to an MkIS, wherein communication between the system and managers is essentially one-way (i.e., in the form of regular reports from the system to the managers), an MDSS allows managers to *interact* with it and make nonroutine requests for information. For example, a product manager may want to know what will happen to sales of brand X in a certain region if its price is cut by 10%. A well-designed MDSS will be able to respond quickly to such a query.

Expert systems (ESs) are currently at the cutting edge of information systems being developed to aid managers. An ES, in addition to responding to questions posed by managers, can proactively make them aware of market situations guarantee warranting their attention and recommend appropriate courses of action. For example, an ES employed by a grocery store, upon sensing that the store's allocation of shelf space to different cereal brands is suboptimal, would alert management about this situation and suggest how shelf space should be reallocated to maximize profits from cereal sales. Expert systems use artificial-intelligence technology that attempts to imitate the mental capabilities and reasoning processes of human experts.

Even though marketing applications of expert systems, the latest stage in the evolution depicted in Exhibit 2.1, already exist, few firms employ such systems for making marketing decisions at the present time. In fact, many firms, particularly smaller ones, still rely solely on *ad hoc* marketing research projects (the very first stage in the evolution) rather than marketing information systems for obtaining information about their markets. According to a recent study, even among the *Fortune* 1000 firms, only 76% have marketing information systems.[3] The upshot is that the evolution of systems supporting decision making is by no means complete. It is likely to continue at an accelerated pace as more and more firms take advantage of existing and emerging information technologies to upgrade their systems.

MARKETING INFORMATION SYSTEM

As implied in the preceding section, a firm that obtains and utilizes marketing inputs on a regular, systematic basis can be said to have a marketing information system. An often-used definition of a marketing information system is given next.

Määritelmä:

Definition. A ***marketing information system*** is a continuing and interacting structure of people, equipment, and procedures designed to gather, sort, analyze, evaluate, and distribute pertinent, timely, and accurate information to marketing decision makers.[4]

asiallista ajankohtaista exact

Notice that the key tasks involved in conducting marketing research—data collection, recording, analysis, and interpretation—are implied in this definition. Thus marketing research techniques and principles are also relevant and vital for the proper functioning of a marketing information system. The main distinction between a marketing research project and a marketing information system is that the latter is a regular source of routine marketing reports (e.g., sales forecasts, market share analyses). A specific marketing research project, in contrast, is intended to meet nonroutine information needs that cannot be met adequately by a firm's marketing information system. But since specific marketing research projects and marketing information systems are both intended to generate information, the tasks associated with obtaining data and converting them to information are germane to both.

tulkinta

liittyvät / koskevat

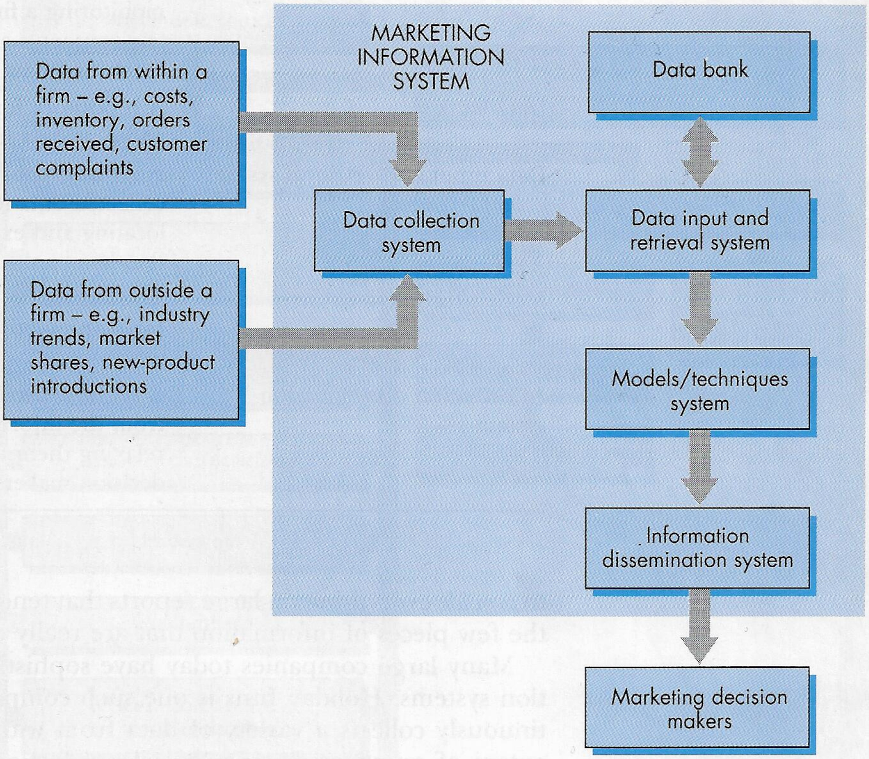
Exhibit 2.2 shows the basic structure of a marketing information system, and Table 2.1 briefly describes its key components. The effectiveness of a marketing information system critically depends on how well

---

[3] Raymond McLeod, Jr., and John C. Rogers, "Marketing Information Systems: Their Current Status in *Fortune* 1000 Companies," *Journal of Management Information Systems*, Spring 1985, pp. 57–75.

[4] Adapted from Philip Kotler, *Principles of Marketing* (Englewood Cliffs, N.J.: Prentice-Hall, 1983), p. 55.

Exhibit 2.2 *Structure of a Marketing Information System*


MARKETING INFORMATION SYSTEM
Data from within a firm – e.g., costs, inventory, orders received, customer complaints
Data from outside a firm – e.g., industry trends, market shares, new-product introductions
Data collection system
Data bank
Data input and retrieval system
Models/techniques system
Information dissemination system
Marketing decision makers

the components are designed and interlinked.[5] Just as good interaction between researchers and research users is essential for the ultimate usefulness of a marketing research project, good communication between designers and potential users of a marketing information system is essential for the effectiveness of the system. Moreover, designers of information systems have the delicate and difficult task of accommodating requests from multiple users and yet ensuring that the system does not become too complex to be useful. As two authors with expertise in information systems have observed, "Some executives play safe and ask for everything. In response to these demands, some designers have tried

---

[5] Design and implementation aspects of marketing information systems are discussed in several books on the subject, such as Van Mayros and D. Michael Werner, *Marketing Information Systems—Design and Applications for Marketers* (Radnor, Pa.: Chilton, 1982); see also Van Mayros and Dennis J. Dolan, "Hefting the Data Load: How to Design the MkIS That Works for You," *Business Marketing,* March 1988, pp. 47–69.

Table 2.1 *Description of Marketing Information System Components*

Data collection system	A set of capabilities for continuously monitoring a firm's internal and external environments and extracting relevant marketing data from them
Data bank	A repository for data collected from within and outside a firm
Data input and retrieval system	A set of capabilities for transferring the collected data to the data bank and for locating and extracting needed data from the data bank
Models/techniques system	A set of frameworks, procedures, and tools for manipulating data to obtain useful information
Information dissemination system	A set of capabilities for generating reports from the information obtained and relaying them to appropriate marketing decision makers

to *provide* everything on large reports that tend to hide rather than reveal the few pieces of information that are really required."[6]

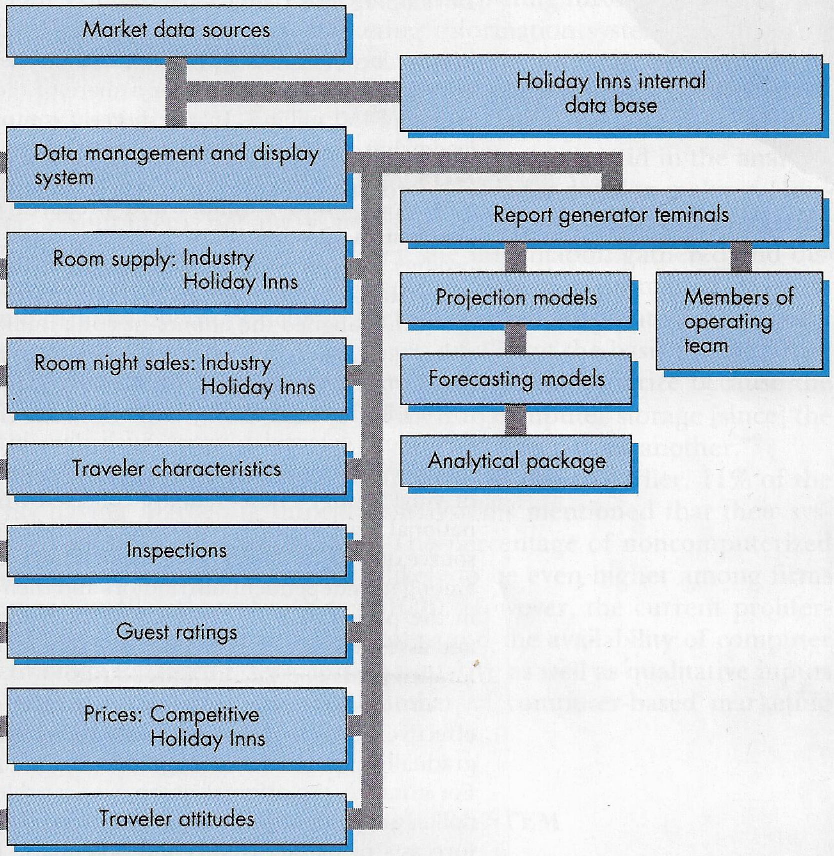
Many large companies today have sophisticated marketing information systems. Holiday Inns is one such company. The hotel chain continuously collects a variety of data from within (e.g., occupancy rates, extent of guest satisfaction and dissatisfaction) and from outside (e.g., government and industry data on total lodging supply, demand for lodging). The collected data are regularly analyzed to generate information that management uses to make effective decisions in such areas as selecting markets for new hotels, allocating advertising spending geographically, tracking market shares, and setting prices.[7] An outline of Holiday Inns' marketing information system is shown in Exhibit 2.3. Even though the structure of Holiday Inns' MkIS does not exactly match the general MkIS structure displayed in Exhibit 2.2, the similarity between the *basic components* of the two structures is evident. Relevant data for Holiday Inns' MkIS are obtained from within the company (i.e., "Holiday Inns internal data base") as well as from outside (i.e., "Market data sources"). These data are assembled, analyzed, and channeled through "Report generator terminals" and the "Data management and

---

[6] Joseph St. John Bate and Dinesh B. Vadhia, *Fourth-Generation Languages Under DOS and UNIX* (Oxford, England: BSP Professional Books, 1987), p. 8.

[7] Thayer C. Taylor, "Computers That Plan," *Sales & Marketing Management,* December 7, 1981, pp. 42–47.

Exhibit 2.3 *Marketing Information System at Holiday Inns*


Market data sources
Holiday Inns internal data base
Data management and display system
Report generator teminals
Room supply: Industry Holiday Inns
Projection models
Members of operating team
Room night sales: Industry Holiday Inns
Forecasting models
Analytical package
Traveler characteristics
Inspections
Guest ratings
Prices: Competitive Holiday Inns
Traveler attitudes

Source: Adapted from Thayer C. Taylor, "Computers That Plan," *Sales & Marketing Management,* Magazine, Copyright December 7, 1981, p. 44.

display system" (which are analogous to a combination of the "Data collection system" and "Data input and retrieval system" in Exhibit 2.2). The three boxes labeled "Projection models," "Forecasting models," and "Analytical package" in Holiday Inns' MkIS together constitute the "Models/techniques system." As this comparison of Exhibits 2.2 and 2.3 illustrates, while the specific elements making up a firm's MkIS may be unique to the company, they will generally fall under the basic MkIS components shown in Exhibit 2.2 and defined in Table 2.1.

The use of sophisticated marketing information systems is not limited to large firms such as Holiday Inns. Smaller firms also can and do employ

Research in Use 2.1 *Mastic Corporation*

Developing and Using an MkIS

Mastic Corporation is a producer and marketer of vinyl siding in a market crowded by 17 firms vying for a share of the total market estimated to be around $450 million. In this fiercely competitive market Mastic holds a leadership position with a 16% share, partly because of a computer-based MkIS.

Mastic's MkIS monitors and prepares regular management reports about such developments as demographic trends, market shares, and product sales performance by territory. Reports generated by the MkIS allow Mastic to fine-tune its marketing strategies. For instance, one year the firm realigned the boundaries of a number of sales territories (to provide better sales coverage) and added several new territories (to better exploit market potential).

In addition to accessing internal records, Mastic's MkIS obtains relevant data from external sources. Such data include unemployment figures, number of housing starts, changes in housing styles, and the acceptability of vinyl siding. These data are collected from the county level up to the national level and then converted to the sales territory level. Yet another source of data for Mastic's MkIS are regular field surveys the firm conducts among its independent distributors and their dealers. The survey data help in the preparation of reports that rank competitor strength by territory and assess the strength of other distribution channels such as independent lumber yards and factory warehouses.

An important feature of Mastic's MkIS which has contributed to its effectiveness is that it was initially designed as a simple system and then gradually expanded and upgraded as the firm gained experience with it. For instance, the original system was capable of merely summarizing sales dollar and volume figures by territory. It was later upgraded to make forecasts of market trends and potentials. Three years after the original MkIS came into existence, and after several rounds of upgrading and expansion, the system was still evolving. To illustrate, Mastic was expanding the system to forecast an export opportunity index. Data for generating the index would include prices of foreign siding products, the value of the U.S. dollar versus other currencies, and the differences in tariffs between the U.S. and other countries.

Source: This illustration is based on information contained in Tom Eisenhart, "Mastic: Widening the Leadership Gap," *Business Marketing,* May 1988, pp. 51–52.

such systems. A case in point is Mastic Corporation, a producer of vinyl siding, whose MkIS and its role in the firm's decision making are described in Research in Use 2.1. A noteworthy feature of Mastic's MkIS is the evolutionary approach used in designing and expanding it. In general, designing a simple, flexible MkIS and then expanding it gradually as one gains experience with it is preferable to starting out with a complex, rigidly designed system.[8]

saavuttaa parempi monimutk. jäykkä

One common misconception about marketing information systems is that a firm cannot have a marketing information system if it does not have access to a computer. Of course, a computer can be useful in efficiently and effectively operating a marketing information system. It can facilitate storage and rapid retrieval of data gathered from a firm's marketing environment, for instance, and it can also aid in the analyses of the data that generate information relevant for decision making. However, a computer is not always necessary for the operation of a marketing information system, especially when the information gathered and disseminated is qualitative. For example, information on market developments and competitors' activities sent to a firm's corporate headquarters on a regular basis by field salespeople may form the basis for the firm's MkIS. But such an MkIS may be difficult to computerize because the information "does not readily lend itself to computer storage [since] the format for one message or record tends to differ from another."[9]

suor.kyk. tehokas helpottaa palauttaa on as. kuuluva

In the study of the *Fortune* 1000 firms alluded to earlier, 11% of the firms having marketing information systems mentioned that their systems were not computer based.[10] The percentage of noncomputerized marketing information systems is likely to be even higher among firms that are smaller than the *Fortune* 1000. However, the current proliferation of rather inexpensive computers and the availability of computer technology to store and process quantitative as well as qualitative inputs should result in an increasing number of computer-based marketing information systems.

MARKETING DECISION SUPPORT SYSTEM

Marketing information systems are evolving into marketing decision support systems in many firms. A marketing decision support system has greater capabilities than a marketing information system, as the following definition implies.

Definition. A ***marketing decision support system*** is a marketing information system that allows decision makers direct access to it and answers "What if?" questions raised by them.

---

[8] See St. John Bate and Vadhia, *Fourth-Generation Languages Under DOS and UNIX;* see also Leonard M. Field, *Monitoring the Competition—Find Out What's Really Going On Over There* (New York, N.Y.: Wiley, 1988).

[9] Raymond McLeod, Jr., and John Rogers, "Marketing Information Systems: Uses in the *Fortune* 500," *California Management Review,* Fall 1982, p. 109.

[10] McLeod, Jr., and Rogers, "Marketing Information Systems: Their Current Status in *Fortune* 1000 Companies."

Distinguishing Features of an MDSS

Marketing decision support systems differ from traditional marketing information systems in several ways. First, a marketing decision support system has interactive capability since it permits managers to query the system directly. In a firm that has only a marketing information system, managers' requests for special reports have to be channeled through systems personnel who act as intermediaries between the system and the managers.

Second, a marketing decision support system offers almost instantaneous response to managers' requests for information. Managers in firms that have a marketing decision support system can interact with the system through personal computer terminals in their own offices. Therefore answers to questions they might pose to the system are literally at their fingertips. To illustrate, at Colgate-Palmolive, one of a growing number of firms with marketing decision support systems, even regional marketing managers and field sales personnel who are constantly on the road can readily interact with the firm's marketing decision support system. They interact with the system via lap-top computers. This ready access provides current information in detail sufficient to assist them in making sound decisions regarding issues such as allocating shelf space to different products, deciding on the product mix and sizes to offer, and setting prices.[11]

Third, the ability of marketing decision support systems to make predictions under alternative "What if?" scenarios painted by managers enables managers to be proactive and make better decisions. In contrast, managers assisted by traditional marketing information systems can, for the most part, only react to routine reports made available by the systems. Thus, while both types of systems provide inputs for decision making, the inputs from marketing decision support systems are much richer and stimulate more effective responses to market developments. Quaker Oats, marketer of such well-known brands as Gatorade and Rice-A-Roni, owes its leadership market position, at least in part, to its sophisticated marketing decision support system.[12] Research in Use 2.2 describes Quaker's marketing decision support system and its applications. Other firms using marketing decision support systems include American Airlines (for price and route selection decisions), Frito-Lay, Inc. (for price, advertising, and promotion decisions), and Zale Corporation (for evaluating potential store sites).[13]

---

[11] Thayer C. Taylor, "Sales Support Systems in the Computer Age," *Sales & Marketing Management,* March 1989, pp. 72–73.

[12] Trish Baumann, "How Quaker Oats Transforms Information into Leadership," *Sales & Marketing Management,* June 1989, pp. 79–80.

[13] Jack T. Hogue and Hugh J. Watson, "Management's Role in the Approval and Administration of Decision Support Systems," *MIS Quarterly,* June 1983, Table 1, p. 18.

Research in Use 2.2 *Quaker Oats*

An MDSS and Its Applications

Quaker Oats, headquartered in Chicago, has been a pioneer in developing and effectively using marketing information and decision support systems. Even as early as 1974, when the concept of a marketing decision support system was foreign to most companies, Quaker had developed its own computer software program, called Addata, that enabled its brand managers to interact with its marketing information base. The resulting marketing decision support system, which has been updated and upgraded constantly since it was first set up, is currently equipped to process and extract insights from about 2 billion facts pertaining to its own products and those of its competitors. Effective use of this system by Quaker's managers has been credited with the firm's achieving and maintaining the highest market share for such brands as Quaker Oats cereals, Gatorade, Van Camp Pork & Beans, Rice-A-Roni, and Aunt Jemima Pancakes.

The importance given to marketing decision support systems at Quaker Oats is reflected by the fact that the company has established a "Marketing Decision Support Services Group" to manage and upgrade the system, and to assist the system's users. The group is headed by the Director of Decision Support Systems. Quaker's MDSS now uses a commercially available software package called Express MDB, which has much better capabilities than Quaker's original Addata package. Express MDB is also very user-friendly with helpful menus. Because of the system's ease of use, even managers new to the system can start using it after just a few hours of training.

According to Steve Rubinov, Director of Decision Support Systems, "[The usage of Quaker's MDSS] falls into three categories. First, reporting and tracking, which includes running the standard reports. Second, marketing planning, where we automate the brand planning and budgeting process by adding 'what-if' analysis and marketing capabilities. Third, ad hoc queries, where people get immediate answers to spontaneous marketing-based questions." The first application is similar to what a traditional MkIS is designed to do, while the second and third applications reflect unique capabilities of an MDSS.

Quaker's MDSS is accessed by over 400 decision makers daily. The following comments from several users illustrate specific applications of the system.

Nancy Bydalek, brand manager for Quaker's Van Camp products: "By running what-if scenarios and marketing spreadsheets based on such considerations as forecasted volume, prices, and advertising spending, I get a national view of my business compared to the competition, so I can identify geographical areas that are doing well and not so well."

Greg Peterson, marketing manager of the Cornmeal brand: "When we

(Continued)

Research in Use 2.2 *Continued*

plan a specific promotion, we go back in time and see the bottom-line effect that different promotional events had on sales. We then plug in the cost of a planned promotion and see what the final effect is going to be on the brand's volume and profit."

Don Scheibenreis, marketing associate for Oh!s cereal: "These reports [on brand performance] help us remain competitive by better managing our financial position. They range from variance reports that specify percentage change for each line item when comparing one financial plan with another to expense reports that detail merchandising costs by month and account."

Source: This illustration is based on information contained in Trish Baumann, "How Quaker Oats Transforms Information into Leadership," *Sales & Marketing Management,* June 1989, pp. 79–80.

Fourth-Generation Languages

As implied by the foregoing discussion, a distinguishing feature of a marketing decision support system is the direct interaction that is possible between the system and its users. An important recent development that is accelerating the pace at which marketing information systems are being upgraded to marketing decision support systems is the availability of computer programming languages called fourth-generation languages (4GLs).[14] A ***fourth-generation language*** is a computer software package that allows even individuals with no training in computer programming to directly access and manipulate data bases. An example of fourth-generation languages is Express MDB, the software package employed in Quaker's marketing decision support system (Research in Use 2.2).

Fourth-Generation Language *A computer software package that allows even individuals with no training in computer programming to directly access and manipulate data bases.*

A variety of fourth-generation languages are commercially available today. Most of them have the following features:[15]

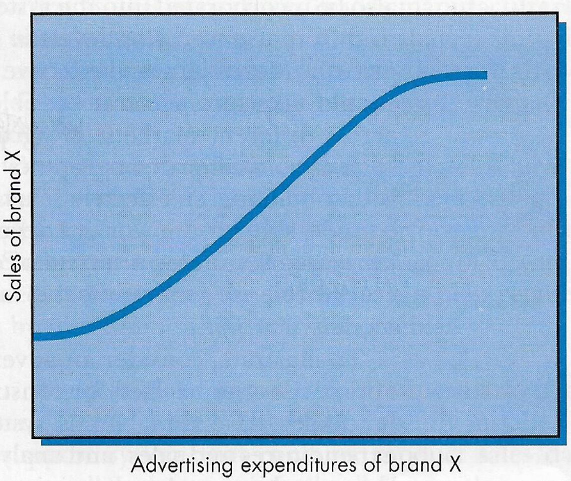
- English-like commands.
- *Instructions accepted by computer in any sequence.*
- High-level commands that the computer expands into detailed technical instructions.
- User-friendliness.
- Capabilities to query data bases for selected items.

---

[14] For a detailed discussion of these languages and their capabilities, see St. John Bate and Vadhia, *Fourth-Generation Languages Under DOS and UNIX.*

[15] List of features is taken from William L. Fuerst, "End-User Computing: A Strategy for the 1990s," *Texas A&M Business Forum,* Fall 1988, p. 24.

Exhibit 2.4 *An Illustrative Advertising-Sales Response Function*



- Easy-to-use requests for reports.
- Graphics capabilities.

These features enable users, regardless of their computer skills, to draw upon their business skills and knowledge and custom design applications to fit their specific requirements.

Marketing Response Functions

The ability to respond to "What if?" questions posed by users, another distinguishing feature of marketing decision support systems, requires such systems to have the necessary models for making the predictions. Therefore incorporating a fourth-generation language alone is not sufficient for upgrading a marketing information system to an effective marketing decision support system. To deliver its full potential a marketing decision support system must also have appropriate marketing response functions built into it. A ***marketing response function*** is a mathematical model that represents the relationship between marketing input and output variables. Exhibit 2.4 contains an illustrative response function that shows how a brand's sales (an output variable) may change in response to different levels of advertising (an input variable).

ennustuksia
riittävä
parantaa

Marketing Response Function *A mathematical model that represents the relationship between marketing input and output variables.*

muuttujia

Response functions such as the one in Exhibit 2.4 have to be constructed for all input variables that managers might include in the "What if?" questions they pose to a marketing decision support system. Input variables other than advertising include price, expenditures on in-store promotion, amount of shelf space allocated, number of sales calls, and so on. The *joint* or *interaction* effects of several marketing inputs must also be incorporated into the system (e.g., the impact on sales and profits of different advertising levels *in combination with* low, medium, or high prices). Moreover, an effective marketing decision support system should take into account possible variations in the response functions across different markets owing to differences in market characteristics such as consumer demographics and levels of competition. The upshot is that building an effective marketing decision support system requires considerable groundwork to develop models capable of simulating market responses under a variety of conditions. Marketing research plays a crucial role in generating the raw material necessary to develop such models.

To illustrate, consider an advertising-sales response function. Several approaches can be used for constructing such a function. One approach is to vary advertising levels systematically, collect data on advertising expenditures and sales, and analyze the data to establish the relationship between the two. An alternative approach is to analyze currently available advertising and sales data from markets in which different levels of advertising were employed in the past. Yet another approach is to ask experienced managers to draw on their market knowledge to estimate the likely sales levels at various advertising levels.[16] The estimates provided by the managers can then be pooled to construct a response function. Because all these approaches involve generating and analyzing data of one form or another, marketing research is clearly an important ingredient in developing marketing decision support systems. Marketing research is also necessary to keep such systems up to date since market characteristics and response functions can and do change over time.[17]

MDSSs with Access to Retail Data Bases

Advances in research methodology and information technology are giving an impetus to the use of increasingly improved marketing decision support systems in many sectors of business. Such systems are perhaps most sophisticated in the retail grocery sector. The advent of the Universal Product Code (UPC) identification system and the availability of automated checkouts with optical scanners are facilitating easy accugathering

---

[16] For guidelines on how to obtain estimates from managers, see, for example, Rakesh Kumar Sarin, "Elicitation of Subjective Probabilities in the Context of Decision-Making," *Decision Sciences,* January 1978, pp. 37–48; Philip Kotler, "A Guide to Gathering Expert Estimates," *Business Horizons,* October 1970, pp. 79–87.

[17] For a good discussion of MDSSs, see John D. C. Little, Lakshmi Mohan, and Antoine Hatoun, "Yanking Knowledge from the Numbers: How Marketing Decision Support Systems Can Work for You," *Industrial Marketing,* March 1982, pp. 46–56; John D. C. Little, "Decision Support Systems for Marketing Managers," *Journal of Marketing,* Summer 1979, pp. 9–26.

mulation of product, market, and sales data in the form of rich data bases. These data bases have greatly increased the capability of supermarket operators, as well as packaged-goods manufacturers, to monitor and analyze sales trends, changes in brand shares, shifts in consumer preferences, and so on. A majority of supermarkets are already equipped with optical scanners and within the next decade virtually all of them will be so equipped.

Retailers whose marketing decision support systems have access to the data generated by computerized optical scanners can get instant and accurate feedback about the relative sales performance of the products they carry. Such continuous feedback, when coupled with the various costs associated with carrying the products, can help retailers maximize their profits by fine-tuning their marketing mixes.

Example. Shop 'n Save is a chain of supermarkets based in the northeastern United States. This chain recently installed a marketing decision support system that contains product sales data from the checkouts, as well as data on virtually all factors that contribute to the cost of the products carried. These factors include shipping expenses, warehouse handling requirements, bulkiness of display, energy needs at the store, time spent on store shelves, promotional expenses, and so on. By accessing these data through the marketing decision support system, managers can perform a direct product profitability (DPP) analysis for each product. The DPP analysis enables managers to pinpoint products that are draining profits and to identify corrective actions. Shop 'n Save's profits-to-sales ratio increased sharply (from 1% to 1.8%) after the computerized decision support system was installed and started being used.[18]

Producers of packaged goods are also benefiting from the wealth of data being generated by computerized checkout scanners. They are doing so by subscribing to and linking their marketing decision support systems with one of several commercially available data bases. An example is the BehaviorScan data base which is maintained by Information Resources, Inc., a pioneer in the growing field of scanner-data research. The BehaviorScan data base has grown both in size and scope since it was first established in the late 1970s.

The heart of BehaviorScan is a central computer located at the firm's headquarters in Chicago. At the present time the computer constantly receives a variety of data from 10 different market areas spread across the United States. These data are transmitted electronically from two sources: a panel of scanner-equipped supermarkets and a panel of households in each of the 10 market areas. BehaviorScan data are available at the individual household level and fall into three categories:

---

[18] "At Today's Supermarket, the Computer Is Doing It All," *Business Week,* August 11, 1986, pp. 64–65.

1. *Household demographics.* Data on such characteristics as household size, number and ages of children, education of household heads, and so on are logged into the central computer when a household joins the BehaviorScan panel.
2. *Supermarket purchases.* Data on what each panel household buys are transmitted from the checkout scanners to the central computer. To tell the computer which household is making the purchases a magnetic card (similar to a credit card) containing the household's identification number is used. This card, carried by the household member making the purchases, is scanned at the checkout along with the products purchased.
3. *Television viewing.* Data on what channels are watched, when, and by whom are gathered electronically by a device (called the peoplemeter) attached to the panel households' television sets. These data are transferred automatically to the central computer.

Access to the BehaviorScan data base can provide packaged-goods marketers with rich insights into a variety of issues such as the relative popularity of brands (in different markets and among different customer groups) and the effectiveness of different types of promotion. In addition to gaining insights from analyzing the data already in the data base, marketers can also use the BehaviorScan system to conduct experiments to determine, for example, which of two alternative television commercials for a product is more effective. The BehaviorScan system has the capability to insert several test commercials within the same television program in such a way that each commercial is beamed to the television sets of a different but representative subsample of panel households. The central computer, in addition to keeping track of the households' purchases of the product in question, knows which specific households received each commercial. Therefore the sales impact of each commercial can be estimated fairly accurately.

Example. The Campbell Soup Company, in attempting to improve sales of its Swanson frozen dinners line, developed two alternative commercials and tested them using the BehaviorScan system. One commercial showed a young man impressing his girlfriend with a fine meal which she believed he had cooked but which was really a Swanson frozen dinner. The other commercial showed a Swanson frozen dinner as a substitute for hot dogs or hamburgers. After experimenting with the two commercials for merely six months, BehaviorScan data clearly showed that the first commercial was significantly more effective—purchases of Swanson frozen dinners among households exposed to this commercial was 20% higher than among households exposed to the second commercial. On the basis of these results the first commercial was rolled out nationally.[19]

Two other research firms that have developed data base systems similar to BehaviorScan are the A. C. Nielsen Company (its system is called ScanTrack) and SAMI-Burke (its system is called ScanAmerica).[20] Data bases such as BehaviorScan, ScanTrack, and ScanAmerica are increasingly being labeled as single-source data bases. A ***single-source data base*** is one that contains, in an integrated fashion, data for individual households on household characteristics, product purchases, and exposure to marketing stimuli. The last type of data (i.e., exposure to marketing stimuli) incorporated into presently available single-source data bases are mostly limited to exposure to television commercials. However, in the near future such data bases will include data on exposure to a much wider range of marketing stimuli aimed at consumers by manufacturers as well as retailers. According to Paul Schmitt, an executive in A. C. Nielsen Company, the company's ScanTrack system will be expanded to capture a full range of data such as those shown pictorially in Exhibit 2.5.[21] Packaged-goods firms whose marketing decision support systems are linked to a data base such as the one in Exhibit 2.5 will have greater ability to make sound decisions concerning a broader array of marketing mix variables.

Single-Source Data Base
Contains, in an integrated fashion, data for individual households on household characteristics, product purchases, and exposure to marketing stimuli.

EXPERT SYSTEM

The vast quantities of data available today are far outstripping the capacity of marketing decision makers to analyze the data efficiently and derive the maximum possible insights from them. This data "explosion," coupled with the emergence of artificial-intelligence technologies, is fueling the development of a new generation of decision support systems in the form of what has come to be known as an expert system.

Definition. An ***expert system*** is a computer-based, artificial-intelligence system that performs tasks normally associated with human intelligence; it is programmed to reason and infer, rather than merely crunch numbers or store and retrieve information.[22]

---

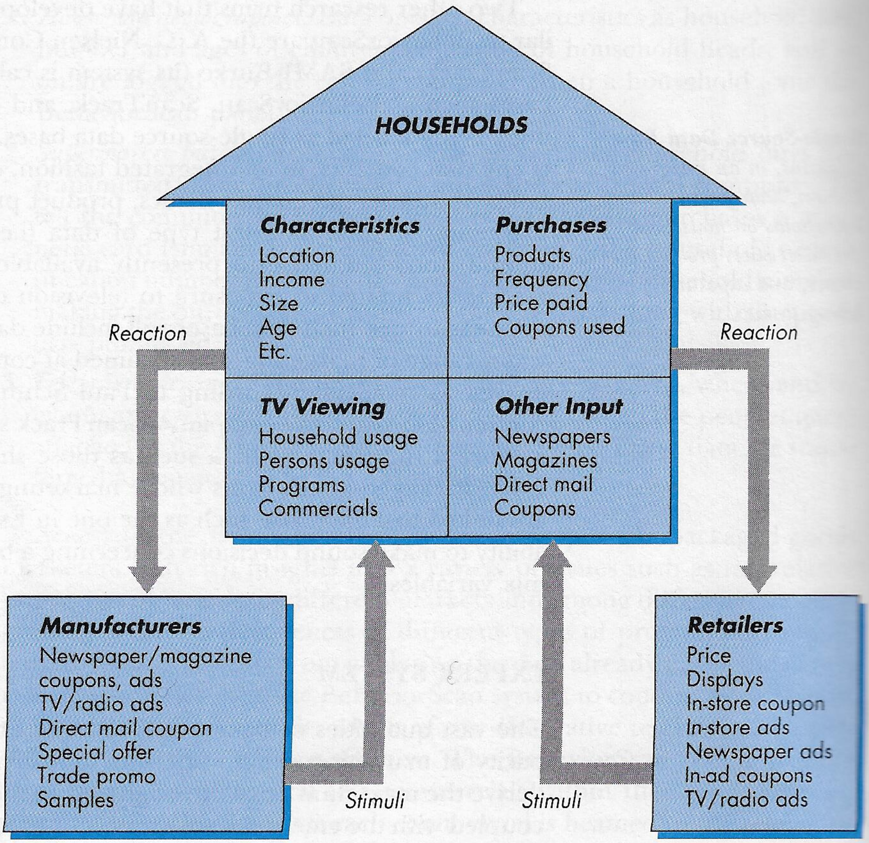
[19] "Big Brother Gets a Job in Market Research," *Business Week,* April 8, 1985, pp. 96–98.

[20] For a discussion of the nature and capabilities of all three systems, see, for example, Felix Kessler, "High-Tech Shocks in Ad Research," *Fortune,* July 7, 1986, pp. 58–62; "Stalking the New Consumer," *Business Week,* August 28, 1989, pp. 54–62.

[21] Paul M. Schmitt, "The Role of Media in ScanTrack," presentation made at the Advertising Research Foundation's Electronic Media & Technology Workshop, New York Hilton Hotel, December 10–11, 1986.

[22] Adapted from Tom Eisenhart, "Texas Instruments: Leave It to the Experts," *Business Marketing,* May 1988, pp. 52–55.

Exhibit 2.5 *Data to Be Captured by Future Single-Source Data Bases*


HOUSEHOLDS
Characteristics
Location
Income
Size
Age
Etc.
Purchases
Products
Frequency
Price paid
Coupons used
TV Viewing
Household usage
Persons usage
Programs
Commercials
Other Input
Newspapers
Magazines
Direct mail
Coupons
Reaction
Reaction
Manufacturers
Newspaper/magazine coupons, ads
TV/radio ads
Direct mail coupon
Special offer
Trade promo
Samples
Stimuli
Stimuli
Retailers
Price
Displays
In-store coupon
In-store ads
Newspaper ads
In-ad coupons
TV/radio ads

Source: Paul M. Schmitt, VP and Director of Business Development, Nielsen Marketing Research, A. C. Nielsen Company, Nielsen Plaza, Northbrook, Ill. 60062. Copyright 1986, A. C. Nielsen Company. Used by permission.

Designing an expert system for use in the marketing field is a greater challenge than setting up a marketing decision support system. In addition to incorporating models and techniques similar to those found in a marketing decision support system, designers of a marketing expert *system have to extract managers' knowledge and expertise* and translate them into a set of decision rules that can be programmed into a computer. *Programming* a computer to reason like a human expert is a formidable task and techniques for accomplishing this task are still in their infancy.

huge

'[illegible]'

Not surprisingly, the use of expert systems in marketing is not yet widespread. However, rudimentary applications of expert systems are

[illegible]

emerging in several areas of marketing. A case in point is the application of expert systems to analyze scanner-based single-source data bases and recommend optimal promotional strategies. According to Douglas F. Haley, Director of Market Information at Nestle Foods Corporation, "Most promising [among analytical tools to process scanner data] are some of the quasi-artificial-intelligence pattern recognition programs. These can extract and analyze promotions using decision rules that duplicate the approaches used by expert analysts: call it a 'brand manager on a chip' if you want to scare your friends in marketing."[23]

Shop 'n Save, the chain store mentioned in the preceding section, uses a form of expert system to determine the best allocation of shelf space among the various products it sells. Shop 'n Save's system, after analyzing sales, cost, and promotional data, translates the results into "Plan-a-Grams," which are "printouts that show store managers, shelf by shelf, exactly where to place their stock to maximize profit."[24] Marketing applications of expert systems are also occurring in companies that do not use scanner-based data. The next example illustrates.

> **Example.** The Balsams Grand Resort Hotel, a 232-room facility located in Dixville Notch, N.H., uses an expert system to make room reservations and assignments. The system is designed to provide the best possible service to guests and, at the same time, maximize room occupancy. It can store up to a million guest histories—demographic characteristics, special needs, preferences, dates and duration of previous stays, and so on. The system holds in memory guest histories for the previous 16 years. When a guest calls to make a reservation, the system quickly suggests rooms that best match the guest's requirements and a final choice can be made with minimum effort on the part of the reservationist and the guest. In addition to improving guest satisfaction, the system has also been very lucrative to the hotel as a result of improved efficiency and higher occupancy rates. Management estimated that the system resulted in additional profits of $50,000 in the very first season after it was implemented.[25]

New artificial-intelligence technologies are continuing to emerge and existing ones are being improved. At the same time, the pressure on marketing decision makers to use their ever-expanding data bases more efficiently and effectively is also increasing. Therefore expert-system applications in marketing are likely to grow rapidly in the coming years.

---

[23] Douglas F. Haley, "Industry Promotion and Advertising Trends—Why Are They Important?" presentation made at the Advertising Research Foundation's Promotion Effectiveness Key Issues Workshop, New York Hilton Hotel, April 22, 1987.

[24] "At Today's Supermarket, the Computer Is Doing It All," p. 65.

[25] Joe Agnew, "Artificial Intelligence with a Marketing Use: Hotel Computer Matches Guests with Their Favorite Rooms," *Marketing News,* September 25, 1987, p. 1.

SUMMARY

A recent trend in the role of marketing research in decision making is an increasing shift from a pure project orientation to a systems orientation. This shift is being driven by a need to use marketing research inputs on a regular, integrated basis. While the trend toward systems-oriented marketing research is still evolving, it has thus far spawned three basic types of systems with increasing levels of sophistication.

The first type of system in the evolutionary sequence is the traditional marketing information system (MkIS). The key building blocks that make up an MkIS include a data collection system, a data bank, a data input and retrieval system, a models/techniques system, and an information dissemination system. The primary role of an MkIS is to keep managers informed of market conditions and developments by generating regular reports and distributing them on a timely basis.

A marketing decision support system (MDSS) is an MkIS that allows managers to interact with it and raise "What if?" questions. The upgrading of MkISs to MDSSs is being facilitated by fourth-generation languages (4GLs), which are user-friendly software packages for directly accessing and manipulating data bases. However, constructing effective MDSSs requires the development of appropriate marketing response functions that show how marketing input variables are likely to influence output variables. Marketing research is necessary to generate such response functions and to keep them current. The application of MDSSs is perhaps most widespread and most advanced in the retail grocery sector, largely due to the availability of single-source data bases such as BehaviorScan which contains a wealth of data about demographics, product purchases, and television viewing behavior of a cross section of households.

The most advanced type of decision-making aid is an expert system (ES). An ES is an upgraded MDSS that makes use of artificial-intelligence technology to imitate the mental and reasoning capabilities of human experts. The use of ESs in marketing is still quite limited but is expected to expand quickly in the near future.

The purpose of Chapters 1 and 2 was to provide an overview of the nature and scope of marketing research, the current status of marketing research, and the role of marketing research in systems for supporting decision making. As demonstrated by this overview, marketing research is an exciting and challenging field, which is growing in sophistication and in terms of potential applications. The appendix to this chapter discusses career opportunities in marketing research.

To give you a flavor of the sophistication with which marketing information is being acquired and used by managers, Chapter 2 focused on the evolving trend toward a systems orientation. However, the basic objective of this book is to discuss procedures and guidelines for conducting research capable of generating useful information. In other words, this is primarily a book on marketing *research* rather than *information systems.* Therefore the remaining chapters will focus on various aspects of marketing research projects. Nevertheless, bear in mind that *ad hoc* research projects are building blocks for constructing sophisticated information systems. For instance, developing response functions, which are essential ingredients of MDSSs, requires conducting appropriate *ad hoc* research projects.

KEY TERMS

Ad hoc research project
Systems-oriented marketing research
Marketing information system
Marketing decision support system
Fourth-generation language
Marketing response function
Single-source data base
Expert system

QUESTIONS

1. Explain the difference between an *ad hoc* research project and systems-oriented marketing research.
2. What two forces are contributing to the move toward systems-oriented marketing research?
3. Define *marketing information system,* and briefly describe the five components that make up the system.
4. Define *marketing decision support system.*
5. State three characteristics of a marketing decision support system that distinguish it from a marketing information system.
6. What is a *fourth-generation language?* Is the availability of an appropriate fourth-generation language sufficient to upgrade an MkIS to an MDSS? Why or why not?
7. What is a *marketing response function?* Briefly describe three approaches for developing response functions.
8. What is a *single-source data base?* Discuss, with suitable examples, several unique benefits to a marketer who has access to a single-source data base.
9. Define *expert system,* and illustrate one marketing application of it.

APPENDIX TO CHAPTER 2

Careers in Marketing Research

The first chapter and the material covered in this chapter presented an overview of what marketing research practice is all about. If this overview has generated some interest in marketing research, you may ask, What are the job opportunities in the marketing research field? However, even if some students get excited about a marketing research *course,* they seldom get interested in thinking about a *career* in marketing research. Unfortunately, the lack of interest in exploring job opportunities in the marketing research field may stem from certain erroneous assumptions on the part of students. The following are typical of such assumptions:

"I have never seen a recruiter visit campus to fill a marketing research job—there must not be any job opportunities out there for me in marketing research."

"I don't get turned on by statistics. I would rather work with people than numbers. So why should I take a job that will keep me busy crunching numbers?"

"If I take a marketing research job, I will be entering a dead-end street—I cannot hope to move on to senior management positions."

The purpose of this appendix is to dispel the myths about marketing research jobs and to present a balanced picture of the opportunities and constraints in a career in marketing research.

MARKETING RESEARCH JOBS ARE AVAILABLE

It is true that few, if any, firms recruit on campus for marketing research positions.[1] However, this fact does not imply that job openings in the marketing research field are scarce. According to research industry sources, the demand for marketing researchers is quite high.[2] Indeed, a recent survey of research firms—research suppliers and field services—found that 74% hire college graduates *fresh from college.*[3] Marketing research positions are also available in numerous client firms with in-house marketing research departments. In addition, a variety of organizations that may not necessarily have formal marketing research departments—such as TV stations, newspaper publishers, banks, nonprofit agencies, or retail institutions—do have exciting research positions.

THE NATURE OF MARKETING RESEARCH JOBS

The jobs available in the marketing research field cover a wide range of tasks. The following jobs are examples (with key tasks involved in each):

- Fieldworker (interviewing).
- Clerical supervisor (overseeing fieldwork).
- Research analyst (designing projects, constructing questionnaires, analyzing results, preparing reports).
- Statisticians (setting up complex sampling plans and analysis procedures).
- Project director (overseeing research projects and research personnel).
- Manager or director of marketing research (overall management of research-related resources and activities).

The positions of fieldworker and clerical supervisor normally require only a high school diploma; the rest usually require at least a bachelor's degree. Salaries for the positions vary, depending on the nature of the position as well as on the qualifications and prior experience of the research personnel. According to a survey conducted by the American Marketing Association, the average 1988 salaries ranged from $17,000 for full-time interviewers to a little over $61,000 for directors of marketing research.[4] Exhibit A.2.1 summarizes the average salaries for a variety of research positions and the extent of growth in those salaries over a five-year period. Of course, as you might expect, entry into some of the

---

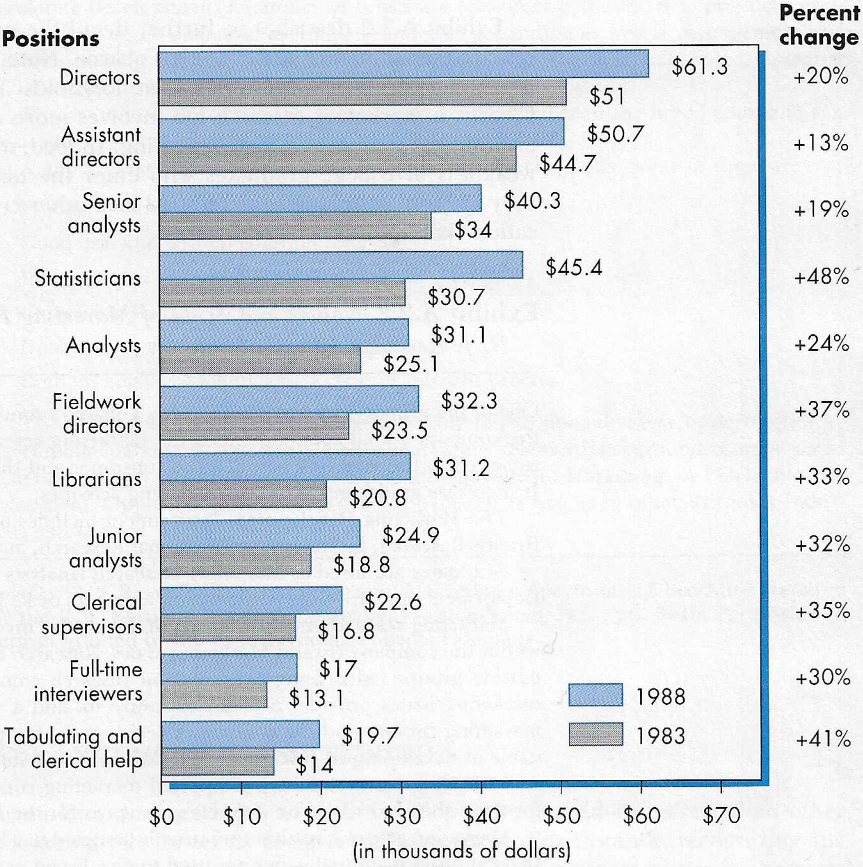
[1] A recent study found that on-campus recruiting was *least used* by firms recruiting for marketing research positions; see A. Parasuraman and Deborah G. Wright, "A Study of Marketing Research Jobs for College Graduates: Implications for Educators," *Proceedings of the American Marketing Association's Educator's Conference, Dearborn, Michigan,* 1983, pp. 181–185.

[2] Jack Honomichl, "Professional Staff Needs Detailed," *Advertising Age,* 52 (October 26, 1981), p. S-34; also see B. C. Vovovich, "Researchers Are Much in Demand," *Advertising Age,* 50 (October 15, 1979), p. S-1.

[3] A. Parasuraman, "Commercial Marketing Researchers' Evaluations of Academic Journals and Marketing Research Education," *Journal of Marketing Education,* Fall 1981, pp. 52–58.

[4] "AMA's 1988 Compensation Survey Divulges Good News and Bad News," *Marketing News,* September 11, 1989, p. 6.

Exhibit A.2.1 *Mean Salaries for Marketing Research positions*



Source: "AMA's 1988 Compensation Survey Divulges Good News and Bad News," *Marketing News,* September 11, 1989, p. 6. Used by permission from the American Marketing Association.

high-salary positions (e.g., director) typically requires a master's degree and a number of years of prior experience in the research field.

Taking a marketing research job is not synonymous with signing a contract to live in a world of numbers and techniques. As the illustrative positions listed earlier imply, working with numbers is only a minor component of many marketing research jobs. Alan R. Cox, group brand research manager for R. J. Reynolds Tobacco Company, had this to say about the nature of marketing research assignments in his company:

> From our viewpoint, research techniques are secondary for marketing researchers who want to do well at R. J. Reynolds. More important is that they have a solid

> marketing background and are, in fact, marketing generalists with research expertise. . . . I do not mean to downplay techniques, but we think a technique is a means to an end (that "end" is solving tough marketing questions) and not the end itself.[5]

Exhibit A.2.2 describes in further detail the nature and scope of marketing research assignments at Reynolds Tobacco. Note especially the types of traits a good marketing research analyst at Reynolds Tobacco is expected to have. Clearly, a marketing research job involves more than, and is likely to be more exciting than, mere number crunching. Indeed, many practitioners feel a major weakness of college graduates who enter the marketing research field is that they are too technique oriented and lack other critical skills, such as communication skills and interpersonal skills.[6]

Exhibit A.2.2 *Nature and Scope of Marketing Research Assignments at R. J. Reynolds Tobacco Company*

One of the primary reasons for Reynolds Tobacco's continued marketing success has been the company's total commitment to the marketing research process. Marketing research is viewed as the vital link between the consumer and the company's marketing activities. It therefore guides much of our marketing activities.

The Marketing Development Department includes both Established Brands and New Brands Research, Strategic and Functional Research, and Marketing Sciences.

Graduates are hired as Marketing Research Analysts into the Marketing Development Department, and are fully participating members of R. J. Reynold's marketing team.

Marketing researchers have thorough day-to-day involvement with other departments within the company (Brand Marketing, Sales, Research and Development) as well as with outside groups (advertising agencies and research companies). This interaction on key marketing issues provides a broad exposure to, and a solid understanding of, both the marketing process and the company's decision-making process. This experience is invaluable in developing R. J. Reynold's marketing researchers into marketing generalists.

Marketing researchers act as internal marketing consultants and are valued as much for their ability to advise on marketing issues as for their research expertise.

Marketing research analysts are usually assigned to a brand as soon as they begin work. Diverse projects on real issues are used to give broad exposure to, and understanding of, key marketing issues. Some of the following areas would be included:

- Advertising and promotion research
- Consumer attitudes and perceptions of brands
- Market segmentation
- Product development
- Strategic planning
- Test marketing

At some time during their first years in the department, marketing researchers are temporarily assigned to different functional areas to help them gain a broader understanding of specific types of research. Examples of functional areas are product research, forecasting, advertising and promotion testing, and sales analysis. These assignments usually last from six to twelve months.

---

[5] Personal communication to the author, dated November 4, 1982.

[6] Parasuraman and Wright, "Study of Marketing Research Jobs."

In addition to this training, company programs and outside seminars are included to accelerate development. Examples of topics are consumer behavior, new product development, advanced analytical techniques, and idea generation as well as management and decision making seminars. Additionally, orientation with the field sales force, Research and Development and advertising agencies are included.

Throughout this program, an analyst is an active member of the brand team and can be a significant contributor to a brand's strategy and tactics.

A good marketing research analyst benefits from certain traits. Some of these are:

- Problem definition and problem solving ability
- Good oral and written communications
- Interpersonal skills and flexibility
- Consulting skills
- Initiative
- Both an analytical ability and a creative thought process

The Marketing Development Department offers exceptional career opportunities by developing marketing generalists with a marketing research background. Career paths generally lead to one of three areas: General Management, Marketing, or Marketing Development. Opportunities exist in Reynolds Tobacco as well as in other Reynolds Industries subsidiaries.

Source: "R. J. Reynolds Tobacco Company: Marketing Opportunities," brochure prepared by the R. J. Reynolds Tobacco Company, Winston-Salem, N.C., pp. 8–10. Permission to reprint granted by the R. J. Reynolds Tobacco Company.

ADVANCEMENT OPPORTUNITIES

A marketing research job is apparently no more a dead-end street than other marketing jobs in terms of career advancement. For instance, recognizing the need to make college students aware of the exciting opportunities in the marketing research field, the Chemical Marketing Research Association (CMRA) has prepared a brochure entitled "Careers in Industrial Marketing Research." It says:

> Experience in marketing research is an effective stepping-stone to management. A recent CMRA study showed that the average non-supervisory marketing research analyst stayed in his job for five to six years. With that background, many analysts progressed to supervisory positions in marketing research and other fields. A substantial percentage rose to become general managers. Others entered the consulting arena by working for management or marketing research consulting firms, while many set out on their own as consultants.[7]

The marketing research jobs at Reynolds Tobacco also have potential for leading to senior management positions, as evidenced by the last paragraph in

---

[7] "Careers in Industrial Marketing Research," a brochure prepared by the Chemical Marketing Research Association, 139 Chestnut Avenue, Staten Island, N.Y. 10305.

Exhibit A.2.2. In many companies a good marketing researcher can move into advertising management and product management, and from there into higher levels of marketing management.

THE CONSTRAINTS

There are a few important, but not necessarily insurmountable, constraints facing a student interested in a marketing research career. First, becoming aware of job openings is not as easy in the marketing research field as it is in fields like sales or retailing. Employers seldom use on-campus recruiting for marketing research positions; rather, they use personal referrals (e.g., from colleagues or college professors) and job advertisements in professional publications (e.g., *Marketing News*) to locate prospective candidates.[8] Consequently, a student interested in a marketing research career must be willing to put out a greater-than-average effort to identify potential employers and to make contact with them.

Second, some real-life experience appears to be a prerequisite for landing a good marketing research job. According to the brochure on marketing research careers prepared by the CMRA:

> Most people enter marketing research only after gaining experience and training in other areas of the company. Sales, marketing, research [and development], production and purchasing are the usual spawning grounds for successful marketing researchers.[9]

Although many marketing research firms (e.g., A. C. Nielsen, Burke Marketing Services) do hire college graduates with no prior experience as trainees, students with some exposure to marketing research are likely to have an advantage over others. The importance of experience is further emphasized by the following types of comments made by potential employers when asked to offer suggestions for students interested in careers in marketing research:

> "Be patient and gain experience—become an excellent listener."
> "Realize that EXPERIENCE is the ultimate teacher."
> "Get as much business experience as you can; volunteer, if necessary."[10]

Hence gaining some experience—for instance, through a business internship sandwiched between blocks of course work—will be a definite advantage.

Third, although marketing research jobs are available to students with only bachelor's degrees, a master's degree is becoming increasingly important, at least in terms of career advancement.[11] An analysis of a sample of ads describing marketing research job openings showed that

> a master's degree was required for over half of all positions that explicitly stated a college degree requirement. Furthermore, a master's degree was preferred for 58% of the positions that required a bachelor's degree. As further evidence of the im-

---

[8] Parasuraman and Wright, "Study of Marketing Research Jobs."

[9] "Careers in Industrial Marketing Research."

[10] Parasuraman and Wright, "Study of Marketing Research Jobs," p. 184.

[11] B. C. Vovovich, "Researchers Are Much in Demand," *Advertising Age,* 50 (October 15, 1979), p. S-1.

> portance of a master's degree in qualifying for the positions advertised, a master of business administration (MBA) degree was mentioned in over one half of the ads that required some specific major.[12]

In view of the importance of experience and a graduate degree for a successful career in marketing research, a few universities have started offering a graduate program in marketing research that has a built-in business experience component.[13]

In conclusion, the marketing research field holds exciting and rewarding career opportunities for interested students, especially those who are willing to expend the extra effort needed to overcome some of the potential constraints we discussed.

SUMMARY

Contrary to the apparent beliefs of many students, the marketing research field does offer exciting job opportunities, involving more than working with numbers and having good career advancement potential. However, for a college student to become aware of job openings is somewhat harder in marketing research than it is in fields such as sales or retailing. Moreover, some business experience and a graduate degree are prerequisites for many marketing research jobs. Nevertheless, these constraints can be overcome by students willing to plan in advance for obtaining employment in the marketing research field.

QUESTIONS

1. Summarize the myths that students typically have about job opportunities in the marketing research field, and indicate why they are not true.
2. Assuming your goal is to start your career in the marketing research field, what specific steps can you take to maximize the chances that your goal will be accomplished?

---

[12] Parasuraman and Wright, "Study of Marketing Research Jobs," p. 182.

[13] See, for example, "Graduate Program in Marketing Research," a brochure prepared by the University of Georgia, Athens, Ga. 30602.

Cases for Part One

CASE 1.1 *ACTION MARKET RESEARCH SERVICES, INC.*

CASE 1.2 *MODERN SILO SYSTEMS*

CASE 1.3 *WALKER RESEARCH, INC.*

CASE 1.1

Action Market Research Services, Inc.

Action Market Research Services, Inc., is a large national firm offering a variety of marketing research and consulting services to client firms in a number of different industries. Among the services it offers is a two-week marketing research training program for client firms' employees. The program is especially designed for individuals who are likely to conduct marketing research on their own and/or use marketing research information frequently in their decision making.

Action Market Research Services devotes the first day of the training program to giving the program participants a feel for the nature and potential usefulness of marketing research. Topics covered on the first day include ways in which marketing research can be helpful in different decision-making situations, what marketing research can or cannot accomplish, and factors crucial for the overall effectiveness of marketing research. In the very first session of the training program, a number of illustrative situations are briefly described to participants, who are then asked to discuss the role that marketing research can play in each situation. Participants are told that they should focus their discussion on the following three issues:

1. Can marketing research be helpful in the situation?
2. If the answer is yes, in what ways and for what aspects of the situation can it be helpful? What types of marketing research activities are likely to be beneficial in generating information that is pertinent to the situation?
3. If the answer is no, explain. What constraints or limitations are likely to prevent marketing research from being truly beneficial in the situation?

Four of the illustrative situations used by Action Market Research Services in its training programs are presented below.

LIFE INSURANCE FIRM

A major life insurance company is becoming increasingly concerned about the growing lapse rate in the life insurance policies it has been selling to customers. *Lapse rate* is defined as the percentage of policy owners canceling their policies in a given year. For instance, if the average number of policies in force during a year is 50,000, and 5000 policy cancellations occur during that year, the lapse rate is 10%. During the past three years this firm's lapse rate has grown from 6% in the first year, to 7% in the second year, to 10% in the third year.

LOCAL LIQUOR STORE

Mr. Dunham, owner of two liquor stores in a city with a population of about one hundred fifty thousand, is planning to open a third store. While the two existing stores are traditional liquor stores, Mr. Dunham wants to make the third store somewhat sophisticated and unconventional. Specifically, he wants to carry only an exotic product line, consisting of fine and rare liquors, wines, cheeses, and breads from all over the world. He plans to attract a socially upscale group of customers to the store.

ELECTRONIC-COMPONENTS MANUFACTURER

A firm producing a line of electronic components has a customer base of about two thousand electronic-goods manufacturers. This firm's research and development (R&D) department has just developed a new design for the component. This newly designed component is expected to greatly increase the reliability of the products into which it is installed. However, the improved version of the component will cost about 30% more than the current version being sold.

PACKAGED GOODS FIRM

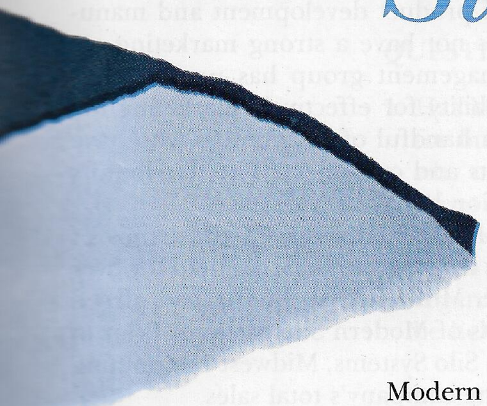
Top management of a firm producing a variety of consumer packaged goods is wondering whether the firm should build another plant to increase its production capacity. The firm's existing plant is operating at 90% capacity, and the firm's sales growth rate has been better than the industry's growth rate during the past three years. Top management feels that the need for increased production capacity depends directly on consumer demand. It therefore decides to ask the firm's marketing research department to investigate the nature of future consumer demand for its products and decide whether a new plant needs to be built at this time.

QUESTION

1. Assuming that you are participating in Action Market Research Services' training program, what are your reactions to each of the four situations presented? Be sure to address the specific issues that participants in the program are asked to discuss.

CASE 1.2

Modern Silo Systems



Modern Silo Systems, a leading manufacturer of grain silos in the Midwest, has just filed suit against one of its former distributors, Midwest Distributing Company. In its lawsuit Modern Silo Systems contends that Midwest Distributing has illegally copied the design of Silotech, one of Modern Silo Systems' most innovative and lucrative products. According to the lawsuit Midwest Distributing had abruptly severed its relationship with Modern Silo Systems about three years before and had since introduced a silo system of its own called Storesafe, a product very similar to Silotech. Top officials of Modern Silo Systems are in the process of gathering information they hope will show that Midwest Distributing's recent actions are unfair and have adversely affected their company's sales and market position.

COMPANY BACKGROUND

Modern Silo Systems has been in business for over 50 years. Its products are sold to a large number of farmers and farm cooperative organizations throughout the midwestern section of the United States. The company enjoys a leadership image in the industry because of the many advances it pioneered in the areas of treating, storing, and preserving harvested grain. Technological expertise and manufacturing capability are the firm's major strengths and the reasons for the respect it commands among customers as well as competitors. The company's current sales are a little over $65 million.

Modern Silo Systems introduced Silotech, a sophisticated, energy-efficient system for drying and storing grain, about eight years ago. The year in which Silotech was introduced saw a sharp increase in the company's sales—from about $30 million to $48 million. Virtually all of this sales increase was due to Silotech. This new product continued to do well in succeeding years until about three years ago, when Midwest Distributing cut off relations with Modern Silo Systems.

MARKETING OF SILOTECH

While being a leader in the areas of product development and manufacturing, Modern Silo Systems does not have a strong marketing organization. The company's top-management group has no executive with a marketing title. The responsibility for effectively marketing the company's products is relegated to a handful of distributors who have thorough knowledge of their markets and customers. The distributors are compensated on a sales commission basis.

The marketing of Silotech was almost completely in the hands of Midwest Distributing Company. This distributor had accounted for over 75% of Silotech's sales. Furthermore, Midwest Distributing had played a crucial role in selling other products of Modern Silo Systems. Prior to severing its relationship with Modern Silo Systems, Midwest Distributing had accounted for roughly 50% of the company's total sales.

EVIDENCE GATHERED BY MODERN SILO SYSTEMS

Modern Silo Systems' total sales reached a peak of $80 million during the year in which Midwest Distributing broke off relations. Total sales in the next two years were $72 million and $68 million, respectively. Sales in the past year were just over $65 million.

Modern Silo System's top management is convinced that Midwest Distributing's introduction of Storesafe is the primary reason for the company's deteriorating sales performance. The company's engineers are sure that Storesafe's design features are very similar, if not identical, to those of Silotech. Furthermore, investigation of the firm's internal accounting records revealed the following trend in Silotech's unit sales during each of the eight years following its introduction. (*Note:* Midwest Distributing broke off its association with Modern Silo Systems at the end of the fifth year).

Year	Silotech Sales (Units)
1	150
2	180
3	234
4	305
5	400
6	219
7	170
8	160

Modern Silo Systems' managers have heard through several of their current distributors and customers that Storesafe's sales are increasing. They have also found out that Midwest Distributing is selling Storesafe at a price that is about 20% to 30% lower than Silotech's price. According to Modern Silo Systems' accounting department, the gross profit margin (i.e., selling price less total cost) for Silotech is roughly 20% of its selling price.

QUESTIONS

1. Has Modern Silo Systems conducted any marketing research relevant for building its case against Midwest Distributing Company? Why or why not?
2. From the information given, do you feel that Modern Silo Systems has a strong case against Midwest Distributing? Explain your answer.
3. What advice would you give to Modern Silo Systems' top management regarding the extent and the types of marketing research that may be helpful in this situation?

CASE 1.3

Walker Research, Inc.

Walker Research, Inc., headquartered in Indianapolis, Indiana, is one of the top 20 marketing research firms in the United States in terms of sales revenue. It has offices in several cities across the country and conducts hundreds of custom research projects for clients every year. Walker Research's management is strongly committed to following sound ethical principles in its numerous dealings with clients, respondents, and other research suppliers. This commitment to research ethics is evident in the following excerpt from an article titled "Marketing Research Ethics Emphasized at Walker," which appeared in the Fall–Winter 1983 issue of *Marketing Researcher,* a quarterly publication of the firm:

> Are you ethical? Most people and organizations think they are. But in most instances ethics are taken for granted and rarely given top consideration. Not so at Walker Research.
>
> Special meetings were recently held with all marketing professionals and managers in Indianapolis, Cedar Knolls, and Phoenix. As a portion of the program, participants reviewed the results of an ethics quiz that was administered to various Walker Research employees. This quiz involved a simulation of various marketing and research-related situations calling for the exercise of judgment and decision making based on ethical and business-related criteria.
>
> To our knowledge, Walker Research is the only organization that holds sessions on ethics to openly discuss such sensitive subjects. However, we believe ethics are an important and vital undercurrent that must be ever-present in the minds of those people working in marketing research.

Source: Information for this case was provided by courtesy of Walker Research, Inc., 3939 Priority Way S. Dr., P.O. Box 80432, Indianapolis, Ind. 46280-0432.

THE ETHICS QUESTIONNAIRE

The ethics quiz mentioned in the preceding excerpt had 15 simulated scenarios covering various aspects of marketing research. Five of those scenarios are provided here.

Scenario 1

A client company is getting ready for labor negotiations and wants you to do a telephone survey among employees to find out how strongly they support union demands. You are given a list of employee names and telephone numbers. If respondents ask why they were chosen to be interviewed, your interviewers are supposed to say that a computer was used to pick a random sample of community residents. What would you do?

a. If respondents ask how they were chosen, tell them, "At random, by a computer."
b. If respondents ask how they were chosen, say you don't know.
c. Insist on telling respondents the truth if they ask or you will not do the survey.
d. Tell the client that you want to add a few other community citizens chosen at random, so you can say that a computer did it.
e. Other.

Scenario 2

You have thoroughly pretested a questionnaire and find that it requires a 45-minute interview. You need a high response rate. The pretest interviewers tell you that if they say in advance that 45 minutes are needed, there are a lot of refusals. But if they say it takes 30 minutes, most respondents will go along and finish the full 45 minutes. What instructions should you give your field interviewers?

a. If respondents ask how long it takes, say 45 minutes.
b. If respondents ask how long it takes, say 30 minutes.
c. If respondents ask how long it takes, say you don't know.
d. Other.

Scenario 3

For a mail survey you are doing, your client wants you to use ultraviolet ink to identify the questionnaires for later classification purposes. Because some of the questions are sensitive and private, the survey itself is announced as anonymous, with no way to connect names with answers. What would you do?

a. Approve the use of ultraviolet ink in order to get the information needed.
b. Do not approve the use of ultraviolet ink.
c. Other.

Scenario 4

For an in-depth interview study among community leaders, your client wants you to have interviewers use tape recorders hidden in their attaché cases. The tape recorders will ensure that you get a record of everything that is said. The client is concerned that respondents might say different things if they knew they were being taped. The client agrees that only you will have control of the tapes and that as soon as you finish the report, you can destroy the tapes. What would you do?

a. Use the hidden tape recorders, and instruct the interviewers to say nothing.
b. Use the hidden tape recorders, but instruct the interviewers to tell the respondents about them if the respondents become suspicious.
c. Do not use hidden tape recorders.
d. Other.

Scenario 5

You have conducted an attitude study for a client. Your data show that the product is not being marketed properly. This finding is not popular with the client's product management, and they request that you omit the data from your formal report on the grounds that the general-overview presentation to management is adequate for management needs. What do you do?

a. Include all data in your formal report.
b. Exclude the marketing data as requested.
c. Other.

For each scenario, the ethics questionnaire also posed the following question to respondents:

Is the above issue mostly a matter of ethics or mostly a matter of research judgment?

a. Ethics.
b. Research judgment.
c. Some of both.
d. Neither.

QUESTIONS

1. What, in your opinion, is the proper thing to do in each of the five scenarios? Why?
2. Is dealing with the issue involved in each scenario simply a matter of research judgment rather than research ethics?

PART TWO

KEY ASPECTS of the Marketing Research PROCESS

CHAPTER 3 *Steps in a Marketing Research Project*

CHAPTER 4 *Types of Marketing Research*

CASES FOR PART TWO

CHAPTER 3

Steps in a Marketing Research Project

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *State and discuss the four criteria for deciding whether a proposed research project should be conducted.*
- *Illustrate why the nature of the uncertainty faced by a decision maker will influence the value of a proposed research project intended to reduce the uncertainty.*
- *Explain with a suitable example why accurate definition of a project's purpose is perhaps the most critical step in executing the project.*
- *List the various key stages involved in a research project and explain the interrelationships among them.*
- *State the circumstances in which execution of a project will not involve all the research stages.*

This chapter will introduce you to the key steps involved in conducting a marketing research project. The objective is to offer a preview of the ***research process***—an interrelated sequence of steps that make up a research project—and, at the same time, lay the foundation for succeeding chapters. At appropriate points during our discussion, specific chapters that cover in detail the elements introduced here will be mentioned.

Research Process
An interrelated sequence of steps that make up a research project.

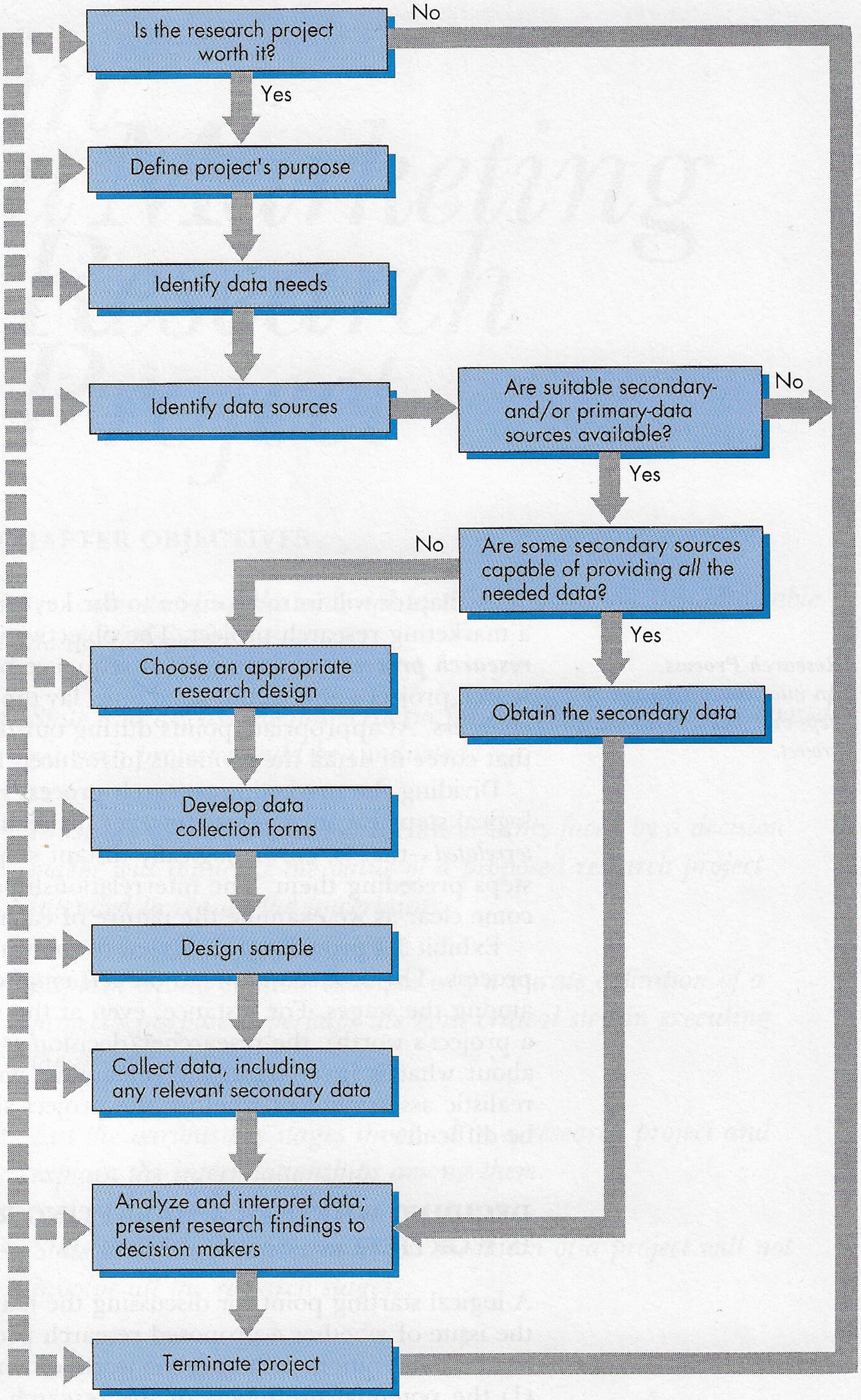
Dividing the marketing research process into a sequence of chronological steps is convenient. However, in reality, the steps are highly *interrelated*—that is, chronologically distant steps may have an impact on steps preceding them. The interrelationships among the steps will become clear as we examine the nature of each.

Exhibit 3.1 provides an overview of the various stages of the research process. The broken arrows to the left emphasize the interrelationships among the stages. For instance, even at the very first stage (evaluating a project's worth), the researcher/decision maker must have some idea about what is involved in the remaining stages. Otherwise, making a realistic assessment of whether the project should be undertaken may be difficult.

DECIDING WHETHER A MARKETING RESEARCH PROJECT IS WORTH IT

A logical starting point for discussing the marketing research process is the issue of whether a proposed research project should be conducted. Four considerations influence the worth of a marketing research project: (1) the potential usefulness of the research results, (2) management's

Exhibit 3.1 *Research Project Stages*


Is the research project worth it?
No
Yes
Define project's purpose
Identify data needs
Identify data sources
Are suitable secondary- and/or primary-data sources available?
No
Yes
No
Are some secondary sources capable of providing *all* the needed data?
Yes
Choose an appropriate research design
Obtain the secondary data
Develop data collection forms
Design sample
Collect data, including any relevant secondary data
Analyze and interpret data; present research findings to decision makers
Terminate project

attitudes toward the research, (3) the resources available for implementation of the research results, and (4) the cost versus the benefits of the research project. These four considerations are not necessarily independent of one another; however, each is important enough to be discussed separately.

Potential Usefulness

The potential usefulness of a research project is the extent to which its findings will be helpful in taking further action in a given situation. Alternatively, it is the extent to which the research findings are likely to reduce a decision maker's uncertainty or provide relevant additional insights in a given situation. Obviously, a proposed project with little potential usefulness should not be undertaken, at least not before appropriate revisions are made. In reality, however, this seemingly simple point is often forgotten. Research projects undertaken hastily, without adequate consideration of their potential usefulness, typically end up generating incomplete information, at best, or totally irrelevant information, at worst.

A major weakness of the consumer survey conducted in the Edsel case (Chapter 1) was that it merely identified how a potential customer segment viewed the new car but failed to uncover specific physical and style features likely to appeal to that segment. The information generated was therefore not complete enough to be beneficial. Had the potential usefulness of the consumer survey been examined more carefully, *before* the survey was conducted, it perhaps would have generated more complete and more relevant data.

In evaluation of the potential usefulness of research findings in a given situation, the important factors are those found in subsequent steps of the research process—such as the nature of the research objectives and the types of data to be generated through the research.

Management Attitudes

Management, or research users in general, must view a proposed research project with an open mind if the project is to be beneficial. Research effort will be wasted unless decision makers are inclined to accept the research results. As we saw in Chapter 1, marketing research conducted for reasons other than aiding decision making is pseudo-research and of no value to an organization. A case in point is Ford Motor Company's brand name research, in which the findings did not recommend the name Edsel. However, Ford management ignored the research findings since it had apparently already made up its mind to name the new car Edsel. Marketing research is not worthwhile when the potential research users have already decided on a course of action or are unlikely to consider research-generated insights with an open mind.

Resources Available for Implementation

Suppose a marketing research project produces useful recommendations, and management is willing to pursue those recommendations. Can we say this research project has been worthwhile? Not necessarily. The research will still be wasted if management lacks the resources (e.g., money, personnel, time) needed to implement the research results. This point is especially relevant for situations in which a firm may be considering marketing research for uncovering market opportunities, as the following example illustrates.

> **Example.** The XYZ Company is a mass producer and marketer of electric can openers. In an attempt to broaden its product assortment, it is considering a marketing research project to determine the nature and size of consumer demand for electric can sealers. However, XYZ currently has only limited idle-plant capacity and is not sure if it has the money to expand capacity. Moreover, it does not possess the technology to produce electric can sealers on a large scale. Clearly, XYZ would be wise not to spend money on marketing research related to electric can sealers before ensuring the availability of financial and technological resources needed to act on the research results, should the results uncover a market opportunity.

Cost Versus Benefits

The three factors discussed so far—potential usefulness, management attitudes, and resources available for implementation—do relate, although indirectly, to the cost-versus-benefits issue. After all, each of these factors addresses the question, Is it worthwhile to conduct marketing research? The purpose of this section, however, is to focus directly on the *monetary* costs and benefits associated with marketing research.

Research costs are much easier to quantify than research benefits. When the tasks involved in a particular research project are known, estimating the total cost of accomplishing them should be relatively easy. However, one cannot meaningfully estimate the cost of a project without knowing what is involved in its various stages. This observation emphasizes a point we made earlier—the steps in the research process are *interrelated.* From a research user's standpoint, estimating the cost of research is easiest if a research supplier is to be used, because the research cost is simply the price quoted by the supplier.

How does one quantify research benefits? There is no standard way for quantifying benefits. However, since the basic purpose of any research is to aid decision makers *by reducing their uncertainty,* a useful starting point is to examine the *nature* of the uncertainty facing a decision maker. Consider the following scenarios:

> **Scenario A.** Brenda Page, product manager for corporation A, is wondering whether she should conduct marketing research to esti-

mate the likely market share of Natuslim, a new, all-natural, diet soft drink. Specifically, she is wondering if she should test-market Natuslim before deciding whether to introduce it on a full scale.

On the basis of production and marketing cost estimates for Natuslim, its break-even sales translates to a 10% share of the diet-drink market. There are 30 brands of diet drinks currently on the market. Three brands have been leading sellers for a long time and enjoy a market share in excess of 20% each, while each of the other brands has a share of 5% or less.

Ms. Page believes Natuslim is as good as any other brand on the market. However, she is *uncertain* about the market share Natuslim may attract if it is introduced. From her experience and understanding of the diet-drink market, she feels Natuslim's share will be anywhere from 2% to 8%, with an outside chance of being over 8%. Should Ms. Page test-market Natuslim before making a decision about full-scale introduction?

Scenario B. Jerry Wilson, advertising manager for corporation B, is *uncertain* about which of two television commercials, X or Y, to use for his company's brand of deodorant. Commercial X employs a fear appeal: In it a man wears a competing brand of deodorant only to be embarrassed later at a party where everyone avoids him, apparently because his deodorant is ineffective. Commercial Y uses a romantic appeal: In it a man wears corporation B's brand of deodorant, goes to a party, and pleasantly discovers that several women are attracted to him, apparently because of the effectiveness of his deodorant.

Mr. Wilson knows commercials using fear appeals have worked extremely well for some products but have adversely affected the sales of others by disgusting potential consumers. He also feels that while romantic commercials like commercial Y have increased the sales of certain products, they can hurt others by lacking believability in the eyes of potential consumers. Should Mr. Wilson conduct marketing research to test the two commercials before deciding which to use for corporation B's deodorant?

Scenarios A and B depict decision makers faced with uncertainty that can be reduced through marketing research. But is marketing research likely to be equally beneficial in both cases? The answer is no, because the *nature* of the uncertainty is different in each scenario.

Let us look at scenario A first. Test-marketing Natuslim will provide a more precise estimate of its potential market share than Brenda Page's current estimate that it will "be anywhere from 2% to 8%." But the *monetary benefit* of reducing this uncertainty will be negligible. The reason is that the break-even share required for Natuslim is 10%, and Ms. Page feels there is only an outside chance its share will be greater than even 8%. In other words, Ms. Page can make a decision *right now* about full-scale introduction of Natuslim. Intuitively, unless corporation A is inclined to take a very heavy risk, Ms. Page may decide not to introduce Natuslim. Regardless of what Ms. Page's decision is, though, it is not likely to change because of any test market outcome.

You are probably wondering, "Won't Ms. Page still benefit by test-marketing Natuslim? For instance, suppose she test-markets Natuslim and finds its share to be over 10% in the test market. Won't this result give her a sound reason to introduce Natuslim on a broader scale?" While this argument has merit, consider this question: How much trust can or *should* Ms. Page place on a test market outcome indicating a greater-than-10% share, *given her prior expectation that there is only an extremely small chance that Natuslim's share will exceed 8%?* There is always room for error in any marketing research project. Consequently, such a test market result may be unusual and not necessarily guarantee a greater-than-10% share after full-scale introduction. In summary, in view of the circumstances in scenario A—especially the *nature* of Ms. Page's uncertainty—the potential benefits of test marketing will be negligible, if not zero. Indeed, even if Ms. Page does decide on full-scale introduction, her decision should be based on other considerations (e.g., corporation A's willingness to take risk) rather than on a favorable test market outcome.

Let us now examine scenario B. The difficulty and the risk of making a decision in the face of Mr. Wilson's uncertainty are much greater than they are for Ms. Page's uncertainty. Choosing the right commercial can ensure the success of corporation B's deodorant, while choosing the wrong one can ensure its failure. More importantly, marketing research is capable of indicating which commercial, X or Y, is most likely to be the right one. Thus the usefulness of marketing research does appear to be greater in scenario B than in scenario A.

Nevertheless, how can Mr. Wilson *quantify* the benefit of marketing research? There is no standard method for quantifying benefits.[1] However, Mr. Wilson may consider the net increase in profits, if the right commercial is chosen, as an indirect estimate of the potential benefits of marketing research. Of course, Mr. Wilson will still have to estimate the net increase in profits, which is not an easy task. On the basis of past experience, as well as the importance of the deodorant to corporation B, Mr. Wilson may be able to make an educated guess about the monetary value of the marketing research information. A comparison of this "guesstimate" with the cost of the research will indicate whether or not the research should be done. A formal mathematical procedure for determining the value of information, along with the assumptions and estimates needed to use the procedure, is described in the appendix to this chapter.

---

[1] Although several conceptual approaches have been suggested in the literature, they all require quantitative inputs that are often difficult to estimate accurately, at least *before* conducting the research project. See James H. Myers and A. Coskun Samli, "Management Control of Marketing Research," *Journal of Marketing Research,* 6 (August 1969), pp. 267–277.

Two important insights emerge from our discussion of scenarios A and B. First, the monetary benefit of marketing research depends heavily on the nature of the uncertainty it is supposed to reduce. Prior to commissioning research, a decision maker should ponder this question: Are the proposed research and the resultant reduction in my uncertainty likely to make any difference in my decision making? The answer to this question was "extremely unlikely" in Brenda Page's case (scenario A) and "extremely likely" in Jerry Wilson's case (scenario B). The monetary benefit of research will be *zero* if the reduction in uncertainty is not likely to affect a decision.

Second, even in situations where marketing research intuitively appears to be beneficial, quantifying the benefits *objectively* is not easy. A decision maker's subjective judgment, and even creativity, will play some role in quantifying the benefits of a proposed research project. Hence viewing cost-versus-benefit comparisons as purely objective or scientific just because they involve numbers is erroneous. The following example illustrates this point.

Example. Some time ago the author had the opportunity to talk to the president of a large seed company, operating in the Midwest, about his company's use of marketing research. He described an instance when his company wanted to explore the demand for its seeds in the Southwest. He contacted a research supplier, who submitted a proposal bearing a $10,000 price tag. The proposal described a well-thought-out and potentially useful research design aimed at answering the company's questions.

Did the company hire the research supplier to conduct the research? No. According to the president, "The proposal was excellent, but I figured it was worth no more than $6000." When he was asked how he quantified the value of the research project as being no more than $6000, his explanation illustrated the creative and subjective, yet logical, way in which a decision maker can attach a dollar value to a research project.

The president's reasoning ran something like this: "I figured that, instead of commissioning the project proposed by the research supplier, I could hire a full-time salesman for six months—enough time to make several calls on all potential customers in the Southwest. The salesman would cost $1000 a month. At the end of six months I would have a pretty good idea about the demand for our seeds in the Southwest. At worst, the salesman would realize zero sales, in which case we would be short $6000; but then we would also know for sure that there was no demand for our seeds in the Southwest. On the other

> hand, if the sales calls turned out to be productive, we would know there was some demand there; moreover, the revenue generated by the salesman's calls would at least partly pay for his compensation. If you were me, would you pay more than $6000 for the proposed research project?"

In this example the company president compared the investment in research with the investment in a temporary salesman and felt that the latter would yield a better return per dollar invested. Of course, an implicit assumption in the president's approach was that the quality and the accuracy of the information from the salesman would be just as good as those of the research information. While that assumption may or may not be correct, the example demonstrates that quantification of research benefits, in practice, is often subjective. Other approaches frequently used by managers include comparing the cost of a project with quantities such as a firm's sales, profits, or marketing budget, and then deciding whether the cost seems too high. We see, then, that a variety of subjective but reasonable methods can be, and are, used for quantifying the potential monetary worth of a research project.

In summary, prior to undertaking any research project, decision makers as well as researchers must carefully decide whether the project is worth it: Is the proposed research likely to lead to potentially useful findings (i.e., is the proposed research complete and comprehensive enough to yield relevant data)? Will the findings be viewed with an open mind? Are adequate resources available to implement the research recommendations effectively? Will the benefits from the research outweigh the costs? Although answering such questions objectively and accurately may be difficult, failure to consider them at the outset will be tantamount to assuming an *unlimited* research budget. And as we saw in Chapter 1, a major concern of researchers is that the budgets allocated to marketing research are extremely limited. Therefore, researchers must use such meager budgets prudently.

DEFINING THE PROJECT'S PURPOSE

Questions crucial to the effectiveness of any research project are, *What* do we want to find out, and *why*? Indeed, our discussion of whether a research project is worth it implicitly assumed that the research purpose was well defined. The costs and potential benefits of research cannot be meaningfully evaluated unless the "what" and "why" aspects of it are clearly established. Consequently, accurate definition of the research problem or research objective is the key to determining whether research should be done and, if so, the nature of the research to be done. According to Robert W. Joselyn, a respected researcher, "Of all the tasks in a marketing research project, none is more vital to the ultimate fulfillment of a client's needs than an accurate and adequate definition of the research problem. . . . All the effort, time, and money spent from this point on will be wasted if the problem is misunderstood and ill-defined.[2] This point is illustrated in the next example.

> **Example.** Pac N' Sac, Inc., producer of a variety of paper containers for consumer products, suffered a sales decline of 15% during the past year. Alarmed by this sharp decline, and wondering whether a change in promotional strategy would help boost sales, Pac N' Sac's marketing manager contacted the marketing research department. The research department proposed a well-designed, comprehensive customer study to evaluate promotional strategies, including the one then being used. The marketing manager reviewed the proposed study, found it to be sound, and okayed it.

Does anything appear to be wrong with the approach taken by the marketing manager and the research department in attempting to arrest Pac N' Sac's sales decline? Yes, the marketing manager apparently *assumed* the problem was Pac N' Sac's promotional strategy, and the research department went along with that assumption. However, if the assumption is incorrect, the research may be valueless. For instance, what if the cause of the sales decline is really customers' perceptions that Pac N' Sac's products are overpriced or have deteriorated in quality? Under these circumstances the proposed research may be able to identify the best promotional strategy among the alternatives being considered; but employing the best promotional strategy may do little to boost Pac N' Sac's sales if promotion is not the cause of the sales decline in the first place.

Alternatively, what if the total *industry sales* of paper containers had declined by 30% during the past year because of stiff competition from new, nonpaper containers? In this situation the decline in industry sales is the real problem, and Pac N' Sac's sales decline is merely a symptom of it. Furthermore, there is probably not much that marketing research can do to improve Pac N' Sac's sales, since its sales performance is actually much better than the performance of the industry as a whole, despite its 15% sales decline. Therefore any marketing research expenditure will most likely be wasted.

An important lesson emerges from the Pac N' Sac example: At least some time and effort must be expended very early in the research process to identify the correct problem to be researched, *if* there is a problem that marketing research can help tackle. In fact, the firm should usually conduct some exploratory research (a topic discussed at length in the next chapter) simply to define the research problem accurately, before designing a project.

---

[2] Robert W. Joselyn, *Designing the Marketing Research Project* (New York: Mason/Charter, 1977), p. 46.

Accurate definition of a project's purpose requires, first, identifying a number of specific issues pertaining to the situation and, second, deciding which of those issues are worth examining further. In the Pac N' Sac scenario, a variety of potential issues or causes are pertinent to the sharp sales decline: ineffective promotion, high prices, low product quality, a downturn in the industry, and so on. Given these alternative potential explanations for the sales decline, focusing research solely on one issue (e.g., promotional effectiveness) is a mistake. At least initially, research should be aimed at pursuing *all* pertinent issues. While this piece of advice sounds like common sense, many research studies focus on wrong problems or problems that are too narrowly defined due to carelessness and/or faulty assumptions on the part of decision makers. Research projects focusing on such ill-defined problems ultimately result in inefficient use of resources or, worse still, erroneous decisions. Research in Use 3.1 provides a case in point.

Research in Use 3.1 *Coca-Cola Company*

Importance of Proper Problem Definition

In April 1985, the Coca-Cola Company announced its decision to replace its traditional Coca-Cola drink with a newly formulated cola labeled New Coke. This announcement was greeted by unanticipated public outrage and demonstrations to force the company to bring back the old formula. The Pepsi-Cola Company took advantage of this negative reaction and aired commercials in which a chagrined old-Coke fan, apparently forced to switch to Pepsi, sadly asked: "Will somebody out there tell me why Coke did it? . . . First they said they were the real thing; then they said they were it." Pepsi-Cola President Roger Enrico poked fun at Coca-Cola's decision by saying: "Clearly this is the Edsel of the '80s. This was a terrible mistake. Coke's got a lemon on its hands, and now they are trying to make lemonade." Faced with increasingly embarrassing and expensive criticisms, the Coca-Cola Company reversed its decision in July 1985 and announced that it would bring back the old Coke under the "Coca-Cola Classic" label.

Did the Coca-Cola Company conduct any marketing research before deciding to replace old Coke with New Coke: It sure did. In fact, the research preceding New Coke's introduction was one of the most expensive and extensive ever. The company spent about $4 million on taste tests comparing New Coke with old Coke and Pepsi. Over 40,000 consumers in various parts of the country took part in the taste tests. This large a sample size was unprecedented in the soft-drink industry wherein typical sample sizes in new-product tests seldom exceeded four or five thousand

(Continued)

Research in Use 3.1 *Continued*

consumers. In the taste tests, 55% of the respondents chose New Coke over old Coke when the brands were not identified; 52% chose New Coke over Pepsi. In tests where the brands were identified, the margins by which New Coke was preferred over old Coke and Pepsi widened by several percentage points. These findings, coupled with the huge samples on which they were based, convinced Coca-Cola executives that replacing old Coke with New Coke would give them a sizable competitive advantage over Pepsi, which was starting to take market share away from Coke in many major markets.

What, then, went wrong? The consensus of industry analysts and experts is that the research commissioned by the company, though quite sophisticated, focused solely on taste preferences because of the apparent assumption that taste was all that mattered in influencing consumer choice. Thus the research, in a sense, examined the "wrong" problem, in that it failed to examine how consumers would react to no longer having the option of buying the traditional Coke. Issues other than taste preferences obviously had a far greater impact on consumers' reactions. This failure, in turn, stemmed from management's apparent eagerness to rejuvenate its 99-year-old brand and replace it with a formulation that *tasted* better than its pesky competitor, Pepsi. Had management been less myopic and considered the broader ramifications of totally withdrawing old Coke, the "problem" addressed by the consumer research would have encompassed more than just taste preferences. Consequently, the millions of dollars spent on the research might have provided the company with more relevant insights and prevented the total withdrawal of old Coke in the first place.

Source: This illustration is based on information contained in Jaclyn Fierman, "How Coke Decided a New Taste Was It," *Fortune,* May 27, 1985, p. 80; John Greenwald, "Coca-Cola's Big Fizzle," *Time,* July 22, 1985, pp. 48–52.

Effective dialogue between decision makers and researchers is very critical for properly diagnosing any situation calling for the use of marketing research. In the Pac N' Sac example the marketing manager assumed a certain type of problem existed, and the research department did not challenge that assumption. The chances of the wrong or a nonexistent problem's being researched are greatly increased when there is no healthy discussion between researchers and decision makers during the problem definition stage.

Effective dialogue is especially important when the purpose of a research project is *to explore opportunities* (e.g., What is the market potential for our product?) rather than to solve a specific problem. The absence of a problem may tempt decision makers and researchers to pay no more than lip service to defining the purpose of a research project. However, failure to establish and agree upon clear-cut research objectives will decrease the effectiveness of the research and may lead to unnecessary friction between the two parties after the research is completed.

IDENTIFYING DATA NEEDS

What data should be collected during a research project? The ease with which this question can be answered depends on how clearly the purpose of the research project has been defined in the previous stage. Identifying specific data needs involves scrutinizing the research purpose and listing the kinds of data required to accomplish that purpose. Consider the following example.

> **Example.** Consolidated Bakeries, marketer of a national brand of snack food items, is interested in conducting marketing research to evaluate its market position relative to its competitors. Specifically, the purpose of the research is as follows: "To determine our current market share and to ascertain whether our relative market position is likely to improve, stay the same, or deteriorate during the next several years." What kinds of data should Consolidated Bakeries collect through the proposed research project?

The data needed to accomplish the first part of Consolidated Bakeries' research purpose—namely, to determine its current market share—are merely its own sales figures and total industry sales figures. Data requirements for fulfilling the less concrete second part of the research purpose—namely, to assess Consolidated Bakeries' future market position—are not as clear-cut. A variety of data may be able to shed light on what the future looks like for Consolidated Bakeries although none of the data is likely to provide a complete and accurate picture of the future. The following data are examples of the type of data Consolidated Bakeries should collect:

- The *brand loyalty* of consumers toward Consolidated Bakeries' products. Has Consolidated Bakeries been consistently enjoying a loyal clientele up to now? If so, its future market position will probably be as good as its current position.
- The *market shares* of Consolidated Bakeries' *competitors* and the *trends in market shares* during the past several years. Does Consolidated Bakeries have any large, strong competitor(s), or is its competition spread across a number of relatively weak firms? In the past, has Consolidated Bakeries' market share been increasing, holding steady, or declining? What about the shares of its competitors? Answers to these questions can provide some clues about what the future may hold for Consolidated Bakeries.
- The *perceptions of consumers* about Consolidated Bakeries' snack foods and about competing brands on dimensions such as quality, price, and availability.

IDENTIFYING DATA SOURCES

After data needs have been listed, the next logical step is to locate sources capable of providing the data. The relative ease or difficulty in locating data sources will depend on the nature of the data desired. Usually, factual data (such as the number of units of a product sold during the past year) can be obtained through *secondary-data sources,* which we will discuss in Chapter 5. For instance, for Consolidated Bakeries, data on current and past market shares can be readily obtained from commercial research firms such as the A. C. Nielsen Company (offering Nielsen Retail Index[3] and ScanTrack[4] data) and Information Resources, Inc. (offering BehaviorScan[5] data). Data that have already been collected and are readily available from sources like these are called ***secondary data.*** Systems such as BehaviorScan can also provide data on changes in brand preferences of families over time. Such data can be helpful, for example, in accomplishing Consolidated Bakeries' research purpose by indicating the extent of brand loyalty.

Secondary Data
Data collected for some purpose other than the research situation.

Not all data may be readily available, however. For instance, consumer perceptions, listed as one of Consolidated Bakeries' data needs, can only be obtained by directly contacting consumers. Such data are called ***primary data.*** More will be said about primary-data collection in Chapters 6 and 7. Suffice it to say here that collecting primary data entails a significant amount of time and effort on the researcher's part.

Primary Data
Data collected specifically for a project.

When the research purpose can be accomplished through secondary data, at least some of the subsequent stages of the research process may not even be relevant. For example, suppose the sole purpose of Consolidated Bakeries' research is to determine its market share and its competitors' market share. This purpose can be accomplished by simply identifying the appropriate secondary-data sources and obtaining the necessary data from them. The research project will conclude at this stage. An important inference emerging from this observation is that *every project will not necessarily involve all the research stages,* as illustrated by the branches on the right-hand side of the flow diagram in Exhibit 3.1.

We have thus far implicitly assumed that researchers will be able to obtain data (either secondary or primary data) to fulfill all the listed data needs. But what if no source is available for the kinds of data needed in a particular project? For instance, suppose the research objective of Consolidated Bakeries is to ascertain consumer tastes and preferences for snack foods 5 to 10 years from now. There probably will not be any good, trustworthy source of data that will be helpful for this objective, since the time frame is too long. Hence the company should abandon the research project at this stage rather than pursue it further; the research will undoubtedly not be worth it. This example also reinforces the notion of interrelated stages in the research process: What may occur in subsequent stages can have a bearing on earlier stages.

---

[3] "Management with the Nielsen Retail Index Systems," a brochure prepared (in 1978) by the A. C. Nielsen Company, Northbrook, Ill.; see pp. 6–7.

[4] "Nielsen National ScanTrack Service," a brochure prepared (in 1982) by the A. C. Nielsen Company, Northbrook, Ill.; see p. 1.

[5] For details on the BehaviorScan system and its capabilities, see "New 'BehaviorScan' System Ties Sales to TV Ads," *Marketing News,* September 21, 1979, p. 7.

CHOOSING AN APPROPRIATE RESEARCH DESIGN

Once the research objectives and the nature of the data to be collected have been determined, the researchers must choose an appropriate research design, which, in turn, will influence the tasks involved in the remainder of the project. For instance, an industrial-product firm wishing to generate some ideas for improving its product line can do so through informal discussions with selected customers and distributors. The research design underlying this data collection is an exploratory design. In contrast, a department store wishing to obtain a demographic profile of its customers can do so through a formal and structured survey of its customers. This data collection approach is traditionally called a descriptive research design. The next chapter discusses different research designs, the circumstances under which they are appropriate, and their implications for data collection and analysis.

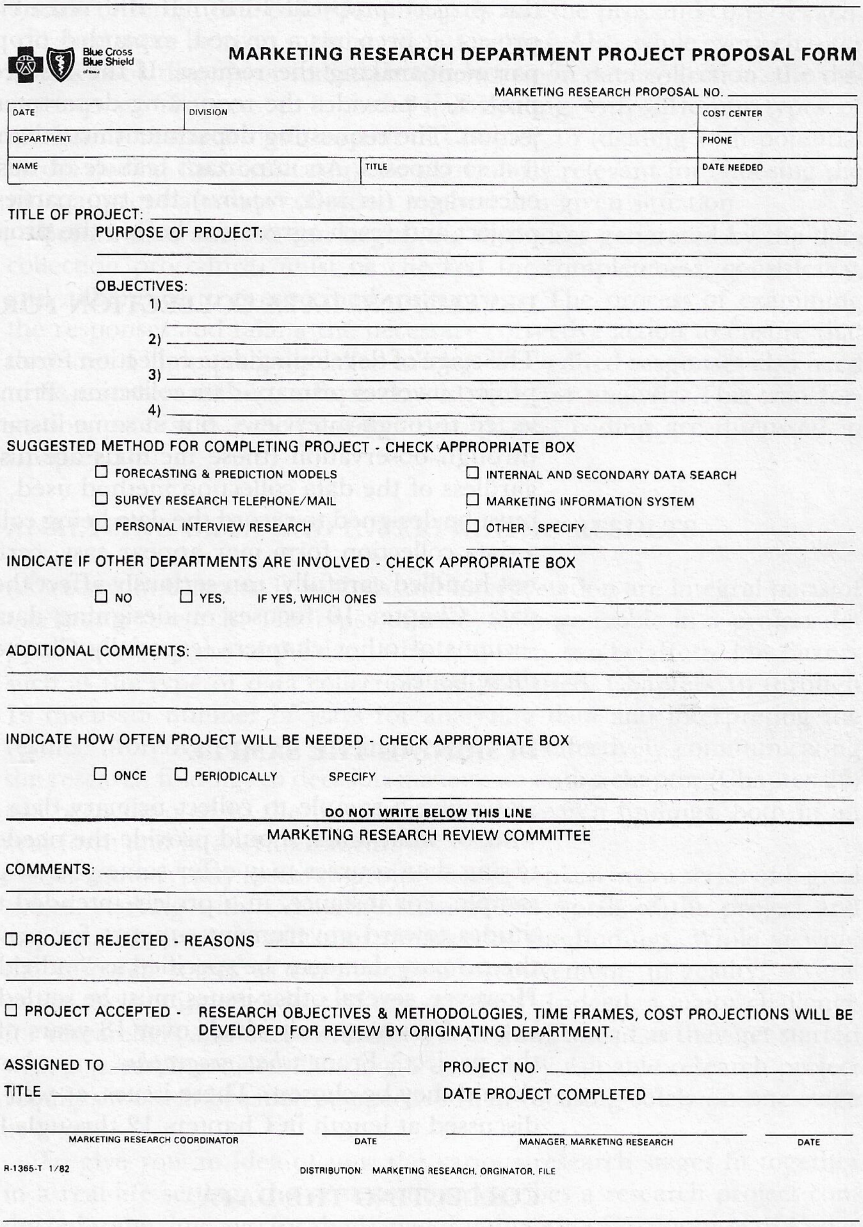
The chosen research design serves as a sort of blueprint for the execution of the project. This blueprint is typically cast in the form of a ***research proposal***—a document that briefly describes the following aspects of the contemplated project: purpose and scope, specific objectives, sample design, data collection procedures, data analysis plan, timetable, and estimated cost. The research proposal can stimulate useful dialogue between the researcher and the decision maker to ensure that there is no misunderstanding between them and that the information to be generated will be appropriate and adequate. As a result of this dialogue, the proposal may have to be modified before a final version that is acceptable to both parties emerges.

Research Proposal
A document that briefly describes the following aspects of the contemplated project: purpose and scope, specific objectives, sample design, data collection procedures, data analysis plan, timetable, and estimated cost.

The initial version of a research proposal is usually drafted by the researcher or research department. However, sometimes the department requesting the research may draft the proposal and submit it to the research department. Irrespective of the origin of the initial proposal, both parties must carefully examine it, raise any concerns they may have, and jointly resolve those concerns by appropriately modifying the proposal.

Exhibit 3.2 contains a standard project proposal form used by Blue Cross and Blue Shield of Iowa. Any department within the company can request that marketing research be done by filling in the top section of this form. This form is essentially the initial draft of the research proposal. After reviewing the request, the research department can

Exhibit 3.2 *Project Proposal Form*



Blue Cross Blue Shield of Iowa

MARKETING RESEARCH DEPARTMENT PROJECT PROPOSAL FORM

MARKETING RESEARCH PROPOSAL NO. ____________

DATE	DIVISION		COST CENTER
DEPARTMENT			PHONE
NAME		TITLE	DATE NEEDED

TITLE OF PROJECT: ____________

PURPOSE OF PROJECT: ____________

OBJECTIVES:
1) ____________
2) ____________
3) ____________
4) ____________

SUGGESTED METHOD FOR COMPLETING PROJECT - CHECK APPROPRIATE BOX

☐ FORECASTING & PREDICTION MODELS
☐ SURVEY RESEARCH TELEPHONE/MAIL
☐ PERSONAL INTERVIEW RESEARCH
☐ INTERNAL AND SECONDARY DATA SEARCH
☐ MARKETING INFORMATION SYSTEM
☐ OTHER - SPECIFY ____________

INDICATE IF OTHER DEPARTMENTS ARE INVOLVED - CHECK APPROPRIATE BOX

☐ NO ☐ YES, IF YES, WHICH ONES ____________

ADDITIONAL COMMENTS: ____________

INDICATE HOW OFTEN PROJECT WILL BE NEEDED - CHECK APPROPRIATE BOX

☐ ONCE ☐ PERIODICALLY SPECIFY ____________

DO NOT WRITE BELOW THIS LINE

MARKETING RESEARCH REVIEW COMMITTEE

COMMENTS: ____________

☐ PROJECT REJECTED - REASONS ____________

☐ PROJECT ACCEPTED - RESEARCH OBJECTIVES & METHODOLOGIES, TIME FRAMES, COST PROJECTIONS WILL BE DEVELOPED FOR REVIEW BY ORIGINATING DEPARTMENT.

ASSIGNED TO ____________ PROJECT NO. ____________
TITLE ____________ DATE PROJECT COMPLETED ____________

MARKETING RESEARCH COORDINATOR ______ DATE ______ MANAGER, MARKETING RESEARCH ______ DATE ______

R-1365-T 1/82 DISTRIBUTION: MARKETING RESEARCH, ORGINATOR, FILE

LAPIN YLIOPISTO KIRJASTO
POISTETTU

Source: T. C. Benedict, Jr., "Flow Diagram Shows How to Evaluate Research Proposals," *Marketing News*, May 1982, p. 11. Reprinted with permission of the author and the American Marketing Association.

either accept or reject the proposed project (see the bottom section of the project proposal form). If the research department accepts the project, it prepares a revised, expanded proposal for review by the department making the request. If the research department rejects the project, it provides the requesting department with reasons for the rejection. The requesting department may then submit a revised proposal if it so chooses. An important feature of this iterative process is that it encourages (in fact, *requires*) the two parties to discuss any proposed project and reach agreement before the project is launched.

DEVELOPING DATA COLLECTION FORMS

The stage of developing data collection forms is relevant when a research project involves primary-data collection. Primary data are frequently collected through interviews, but in some instances they are also gathered through observation (these methods are discussed in Chapter 6). Regardless of the data collection method used, some instrument, or form, must be designed to record the data being collected. Although designing a data collection form may appear easy, certain aspects of the form, if not handled carefully, can seriously affect the quality and nature of the data. Chapter 10 focuses on designing data collection forms; certain sections of other chapters (especially Chapters 6 and 15) also address these issues.

DESIGNING THE SAMPLE

Designing a sample to collect primary data involves clearly specifying who, or what units, should provide the needed data. The stage of identifying data sources may offer some *general* guidance for designing the sample. For instance, in a project intended to measure the public's attitudes toward government support for private schools, the source of the primary data may be specified as "individuals over 18 years of age." However, several other issues must be settled before data collection can begin: *How many* individuals over 18 years of age should be chosen for the project? From *what geographic area* should they be chosen? *How* should they be chosen? These issues, as well as other related issues, are discussed at length in Chapters 12 through 14.

COLLECTING THE DATA

Once the data collection form and the sample design are ready, the next logical step is to collect the data. Note that consistent with the interrelationships among the research process stages, designing the data collection form and the sample should have already taken into account the type of data collection method to be used. For example, the wording to be used in a questionnaire depends on whether it is to be used in a

Editing *The process of examining completed data-collection forms and taking whatever corrective action is needed to ensure that the data are of high quality.*

Coding *A term that broadly refers to the set of all tasks associated with transforming edited questionnaire responses into a form that is ready for analysis.*

personal-interview survey or a mail survey. Chapters 6 and 7, which discuss different data collection methods and the pros and cons of each, will shed more light on such interrelationships. Also, while every chapter in this book directly or indirectly touches upon data collection, the discussions in Chapters 4, 8, and 9 (all dealing with different types of marketing research designs) and in Chapter 15 (dealing with potential errors during data collection) are particularly relevant for choosing the most appropriate data collection method in a given situation.

Before data analysis can begin, the responses generated by the data collection procedures must be checked for completeness, consistency, and adherence to prespecified instructions. The process of examining the responses and taking the necessary corrective action to ensure that they are of high quality is called ***editing.*** The edited responses also need to be transformed into a form that is ready for analysis. This transformation process is called ***coding.*** Editing and coding are discussed in Chapter 16.

ANALYZING DATA AND INTERPRETING RESULTS

As we saw in Chapter 1, analysis and interpretation are integral parts of marketing research. The types of analyses permissible in a project depend on the nature of the data, which, in turn, can be affected by factors such as the type of data collection method used. Chapters 16 through 19 discuss a number of ways for analyzing data and interpreting the results. Moreover, given the importance of effectively communicating the research findings to decision makers, an entire chapter (Chapter 20) has been devoted to the presentation of research findings, both in an oral form and in a written form.

In summary, any research project can be broken into a series of logical stages, starting with a determination of the worth of the project and ending with analysis and interpretation of the findings. While viewing the research stages as chronological is convenient, in reality, several stages are not independent of one another.[6] Indeed, a major challenge for researchers and decision makers is to think ahead as they get started with a research project. Planning a potentially valuable research project requires a much broader perspective than focusing solely on one stage at a time.

To give you an idea of how the various research stages fit together in a real-life setting, the next section describes a research project conducted some time ago for the Wayne Engineering Corporation of Cedar Falls, Iowa.

---

[6] It has even been suggested that, in order to design a project properly, one should *start* by focusing on how the research results will be used (and therefore what the final report should contain) and then work one's way "backwards." See Alan R. Andreasen, " 'Backward' Market Research," *Harvard Business Review,* May-June 1985, pp. 179–182.

A REAL-LIFE ILLUSTRATION

Background

Wayne Engineering is a manufacturer of garbage collection trucks (commonly known as "garbage packers") used primarily by city/town governments and private garbage collection firms. At the time of this study Wayne Engineering had annual sales of around $4 million and was the smallest of 10 garbage packer manufacturers in the United States. Wayne Engineering sold its packers through a national network of dealers. However, it had no sophisticated marketing organization; it had only a few people to coordinate dealer sales activities, plant production, and after-sales service.

Wayne Engineering's sales had remained flat for some time after experiencing rapid growth. According to the company's president, Stan Worthington, the United States market had leveled off for the types of packers produced by the company, which were smaller than the increasingly popular packers produced by competing firms. Consequently, Mr. Worthington wanted to explore the possibility of developing new markets for Wayne's packers. Specifically, he singled out Canada as a new market because of Canada's proximity to the United States (and especially to Iowa, where Wayne's plant was located), the apparent similarities between the United States and Canadian market structures, and the fact that Wayne's United States competitors were already selling a large number of packers in Canada. The author (hereafter referred to as the consultant) worked full-time for Wayne Engineering during the two months when this project was undertaken.

Justification for the Project

A primary reason for considering the project was Mr. Worthington's belief that there might be a market in Canada for Wayne's packers. Several years of relatively flat sales in the United States increased the desirability of conducting the research. Mr. Worthington therefore felt that the research would be potentially useful. Moreover, Wayne Engineering had enough production capacity and other necessary resources to try to tap the Canadian market should the research uncover a favorable market for Wayne's packers. In short, the proposed research appeared to be worthwhile on the dimensions of potential usefulness, management's open-mindedness, and adequacy of resources to implement the research results.

Comparing accurately the *dollar* costs and benefits of the research project was not possible. However, a rough assessment was made along the following lines: The project cost was estimated to be about $6000 on the basis of compensation for the consultant's time and other data collection expenses (such as travel expenses). Estimating the data collection expenses was neither precise nor easy. What the various stages of the project would involve had to be ascertained right at the start of the process—a difficult, but *necessary,* task.

Potential benefits of the project, although worthwhile in a qualitative sense, were impossible to quantify in monetary terms. Therefore an indirect evaluation was made by looking at the size of the break-even packer sales needed to recover the estimated research cost. When the project was proposed, Wayne's packers were selling at an average price of $25,000, and Wayne's contribution margin was 10%, or about $2500 on every packer sold. Thus if only three packers were ultimately sold in Canada as a result of insights revealed by the proposed research, the estimated research cost would be more than recouped. This break-even sales figure was considered reasonable. Also, the estimated cost as a percentage of annual sales was only about 0.15%, a miniscule amount. Hence the research project was deemed to be worthwhile on monetary grounds.

Definition of Research Purpose

The main objective of the research project was to assess potential Canadian customers' reactions to Wayne Engineering's garbage packers vis-à-vis packers already available in Canada. Whether this objective was the *most appropriate* research objective warrants some discussion, especially in view of our earlier emphasis on the need to identify the correct research objective.

The motive prompting Mr. Worthington to consider the Canadian market was the flatness of Wayne's sales over a period of time. However, a key question worth raising is whether options *other than exploring the Canadian market* were even considered. For example, could Wayne Engineering have explored the feasibility of boosting sales through more aggressive selling within the United States or through the introduction of new products? Ideally, issues like these must be jointly discussed by the researcher and the decision maker before deciding on the research purpose. However, in Wayne Engineering's case the consultant did not explore such options because Mr. Worthington (who was more knowledgeable about the company and its markets) had already considered them and dismissed them as being infeasible owing to a variety of constraints. For instance, since Wayne Engineering did not have its own sales force, and since its competition was rapidly increasing in the United States, Mr. Worthington had ruled out the option of trying to increase sales through more aggressive selling. The research purpose as defined was considered to be appropriate, at least in Mr. Worthington's judgment.

Identification of Data Needs

The data requirements for accomplishing the primary research purpose were listed as a set of specific questions:

- What are the main features of the garbage packers used in Canada?
- Are there any similarities or differences between Wayne Engineering's packers and those used in Canada? If so, what are those similarities or differences?
- What type of competition is Wayne Engineering likely to face, from Canadian as well as United States firms, if it decides to enter the Canadian market?
- Are there particular segments in the Canadian market with unique needs capable of being satisfied better by Wayne Engineering's packers than by competing packers? If so, what are the characteristics of those segments?
- What type of Canadian firms or institutions appear to be especially interested in Wayne Engineering's packers? What is the extent of their interest?

Identification of Data Sources

Most of the data to be collected involved potential customers' perceptions about the fit between the Canadian market's needs and the features of Wayne's packers. As a result, some form of primary-data collection was inevitable. Mr. Worthington and the consultant felt that end users of garbage packers, as well as firms involved in distributing and servicing them, would be able to provide data relevant to the study. Specifically, municipalities (i.e., city/town governments), private garbage collection firms, and truck equipment distributors/manufacturers were identified as sources of primary data. In addition, the library and the Canadian Consulate in Chicago were identified as potential sources of secondary data (such as annual unit sales of garbage packers in Canada) pertaining to the research project.

Designing the Questionnaire

Recall that questionnaire design, in addition to considering a research project's objectives and data needs, must also take into account the nature of the respondents and the data collection method to be used (i.e., the stages of designing the sample and collecting the data). Consequently, before designing any questionnaire, Mr. Worthington and the consultant had to think through the procedure to be used in gathering the needed data.

Since Wayne Engineering had little prior data on the Canadian garbage packer market, the project was basically exploratory (more on exploratory research in the next chapter). As is common in many exploratory-research projects, the consultant decided to collect the necessary data through personal interviews of knowledgeable individuals in appropriate Canadian institutions. The personal-interview approach, rather than a telephone or a mail survey, was chosen for the following reason: In order to obtain *meaningful* data, the researcher had to demonstrate the features of Wayne's packers (through pictures and brochures) and to answer any questions the respondents might have. Given the need for such researcher-respondent interaction, telephone and mail surveys were inappropriate (Chapter 6 discusses in greater detail the relative merits and drawbacks of personal interviews vis-à-vis telephone and mail surveys).

The personal interviews were designed to be open-ended and nonstructured because of the exploratory nature of the research; no formal questionnaire was designed and printed. However, the consultant did prepare a set of questions described previously in "Identification of Data Needs." This set of questions was useful in ensuring that all critical topics were discussed during the interviews.

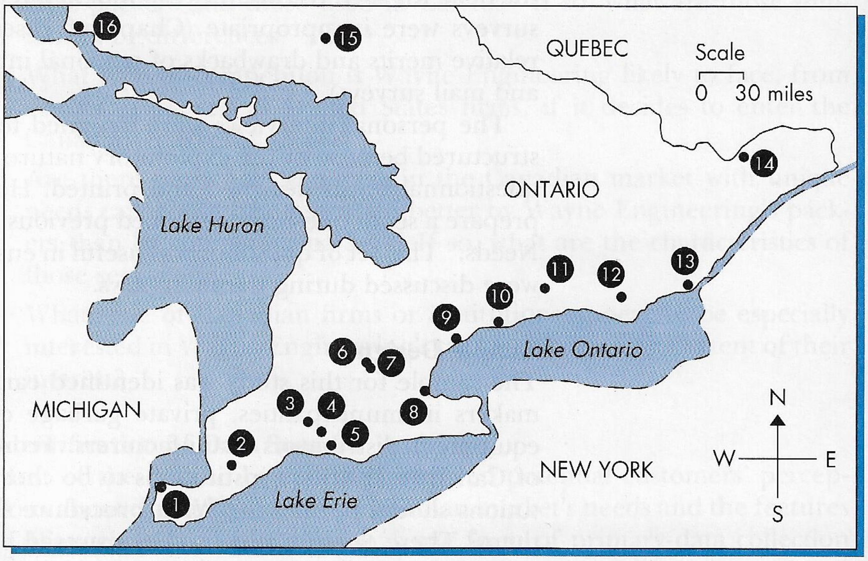
Sample Design

The sample for this study was identified earlier as a group of decision makers in municipalities, private garbage collection firms, and truck equipment distributors/manufacturers. From what geographic area(s) of Canada were these institutions to be chosen? How many such institutions should be chosen? What procedure should be used in choosing them? These questions had to be answered before data collection could begin.

The investigation was restricted to the four Canadian states (or provinces as they are called in Canada) closest to Iowa, namely, Ontario, Manitoba, Saskatchewan, and Alberta. Provinces close to Iowa were chosen because the final price of heavy equipment like garbage packers is very sensitive to the distance between potential markets and production facilities. Markets too far away from Iowa were unlikely to be attractive to Wayne Engineering in terms of price competitiveness. Moreover, the limited time available for completing the entire project precluded broader coverage of Canada. Such a practical constraint is not unique to the Wayne Engineering project; many real-life research studies have to make compromises and use less-than-ideal designs because of limited resources (time, money, personnel, etc.).

Within the four Canadian provinces the consultant further decided to restrict the study sample to cities and towns with populations of 50,000 or more. This decision was based partly on the tight time constraint and partly on the assumption that the demand for garbage packers would be strongest in the major population centers. According to the 1971 census reports the four chosen provinces had over 50% of Canada's population. Within these provinces the consultant pinpointed (with the aid of a road atlas) 21 cities and towns that were convenient to get to during the limited time available for the study. The selected cities together accounted for about 25% of the total Canadian population. This population coverage was adequate given the exploratory nature of the project, although the geographic coverage in terms of the actual number of cities included was relatively small.

Exhibit 3.3 *Consultant's Travel Route Through Ontario*



Key

1	Windsor	5	Aylmer	9	Toronto and suburbs	13	Kingston
2	Chatham	6	Waterloo	10	Oshawa	14	Ottawa
3	London	7	Kitchener	11	Peterborough	15	Sudbury
4	Belmont	8	Hamilton	12	Belleville	16	Sault Ste. Marie

As part of the sample design, the consultant laid out the travel routes to be taken through the chosen provinces. Exhibit 3.3 shows the consultant's travel route through Ontario; the numbers indicate the sequence in which the major cities in Ontario were to be covered. Similar travel plans were drawn up for the other three provinces.

The final sample included 82 institutions. However, owing to the lack of a readily available list, the consultant could not identify all of them by name *before* starting the data collection. Consequently, the institutions were selected judgmentally after the consultant drove into each city on the travel route. Municipal offices in the various cities were obvious candidates for inclusion in the sample, and they could easily be identified and located. However, the yellow pages of the telephone directory had to be used to get the names and addresses of private garbage collection firms and truck equipment distributors/manufacturers.

The yellow pages in relatively small cities contained only a few such firms; hence all of them could be included in the sample. However, in large cities such as Toronto, Ottawa, and Edmonton, the yellow pages listed numerous private garbage collection firms and truck equipment distributors/manufacturers. Since the consultant could not contact all the listed firms, he used the size of the ads placed by the firms as the basis for selecting a few of them. As a rationale, the consultant assumed that firms placing larger ads were likely to be bigger and hence more inclined to be interested in international trade than other firms. Admittedly, this selection process was not very scientific. However, the use of such judgment sampling is not uncommon, especially when a research project is exploratory. (Chapter 14 discusses judgment sampling as well as other subjective, nonprobability sample selection methods.)

Data Collection

Good marketing research practice involves examining potential sources of secondary data before collecting any primary data. Accordingly, the consultant first explored the availability of relevant secondary data in the library and in the Canadian Consulate in Chicago. Unfortunately, neither of these sources had secondary data specific enough to be relevant for the research objectives. However, officials at the Canadian Consulate did offer one important insight: They alerted the consultant to a growing tendency on the part of the Canadian government to restrict the import of foreign goods unless a certain proportion of Canadian content could be added to them. Indeed, this revelation suggested that the study sample should include truck equipment manufacturers, who might be interested in buying garbage packer components or unassembled packers from Wayne Engineering. Moreover, certain types of secondary data were useful at the sample design stage: As we already saw, the Canadian population census data as well as data in a road atlas were helpful in defining the geographic areas to be covered and selecting the specific cities and towns to be studied within those areas.

Primary data were gathered through discussions with decision makers in the selected institutions. At each institution the consultant outlined the purpose of the study and requested a meeting with appropriate personnel. Only 4 of the 82 institutions contacted refused to honor the request for an interview. Most of the respondents interviewed were senior-level decision makers such as purchasing officers, marketing/sales managers, and presidents. Some firms showed a great deal of interest in the survey, and the interviews turned into group discussions with two or more executives participating in them. The length of the interviews ranged from less than 10 minutes to over an hour; on the average, they lasted about 30 minutes.

During each interview, in addition to focusing the discussion on the research questions, the consultant answered any questions the respondents had about Wayne Engineering and also took extensive notes. At the end of each day's round of interviews, the consultant prepared a fairly detailed written report for each institution contacted that day. The interview notes were very useful in preparing these reports. The following details are illustrative of the items included in the reports:

1. Data about the institution, such as its size, nature of operations, and garbage collection products being used or sold.
2. The extent of interest shown by the institution in Wayne Engineering's products and/or in joint ventures with Wayne Engineering.
3. The respondents' perceptions about the Canadian garbage packer market as well as the potential strengths and weaknesses of various garbage packers in the market.
4. The consultant's subjective assessment of the merit of Wayne Engineering's strengthening its association with the institution.

Data Analysis and Interpretation

Because of the exploratory nature of this project and the nonstructured data collection procedure, the data in the written reports were *qualitative* data. Hence quantitative-analysis procedures were irrelevant and inappropriate. The actual analysis performed consisted of examining the written reports to see whether common patterns of response could be identified regarding issues such as the types of packers likely to have the strongest appeal in the Canadian market and the distinguishing features, if any, of firms showing a keen interest in Wayne Engineering. (More will be said about analysis and interpretation of qualitative research data in Chapter 7.)

Several valuable insights emerged from the examination of the written reports. For example, a technologically revolutionary packer for which Wayne had acquired production rights from a Swedish firm generated virtually no interest in the Canadian institutions, much to the surprise of Mr. Worthington. In contrast, one of Wayne's older products, a small packer capable of holding a relatively light load, got an enthusiastic response from many institutions, especially private garbage collection firms. Moreover, 8 out of 24 truck equipment dealers/manufacturers contacted in Ontario were interested in exploring joint-venture or licensing arrangements with Wayne Engineering for producing and marketing packers of this type. In short, the analysis uncovered unique opportunities and constraints for Wayne Engineering.

The consultant made an oral presentation to Mr. Worthington and other Wayne Engineering executives in addition to giving them a copy of the written reports of the interviews. During the presentation the consultant highlighted the types of packers for which there appeared to be good sales potential in Canada, and pinpointed specific Canadian institutions holding the greatest promise for fruitful future relationships with Wayne Engineering.

Deciding Whether the Project Was Worth It

About a year and a half later the consultant contacted Mr. Worthington and asked him whether Wayne Engineering had benefited from the research project. Mr. Worthington felt it had been worthwhile. Several major Canadian cities contacted during the study were in correspondence with Wayne Engineering and frequently requested bids on garbage packers. Moreover, as a result of the study, Wayne Engineering had a list of capable and interested distributors dispersed across the four provinces covered in the study.

Despite these benefits, however, the study had not yet netted any Canadian sales for Wayne Engineering. Mr. Worthington candidly said the study would have been *more beneficial* if they had adequately followed up on the insights and leads generated. But Wayne Engineering had not yet been able to break into the Canadian market aggressively because of personnel and monetary problems. In spite of the earlier effort to weigh the pros and cons of the proposed project, certain potential resource constraints had apparently been glossed over. In retrospect, perhaps Mr. Worthington and the consultant should have more carefully considered questions like the following: What specific action(s) would, or *could,* Wayne Engineering take if the research findings turned out to be such and such?

No doubt the research project offered useful insights about the nature of the Canadian garbage packer market and pointed out several fruitful business contacts in Canada. Nevertheless, had all potential resource constraints been carefully considered at the outset, the research project might have been postponed to a more opportune time when Wayne Engineering could aggressively act on the research findings.

Of course, not all marketing research projects are as unstructured and as informally conducted as the Wayne Engineering study. As mentioned earlier, the nature of the problem and research objectives in a given setting determine the nature of the research design or approach to be used. In subsequent chapters we will see applications of more formal and sophisticated research approaches. Irrespective of the nature of research projects, however, they all involve the same basic interrelated stages, as illustrated by the Wayne Engineering study.

SUMMARY

This chapter introduced you to the various stages involved in planning and conducting a marketing research project. Certain aspects of these stages deserve particular attention. First, although a research project can be viewed as a sequence of chronological stages, the stages are not necessarily independent of one another. Chronologically distant stages may have an impact on preceding stages.

Second, some projects may not involve all the stages shown in Exhibit 3.1. When all the necessary data are available through secondary sources, there is no need to develop a data collection form, design a sample, or collect primary data. At the other extreme, when no trustworthy data source (primary or secondary) capable of adequately meeting a research project's data needs is available, terminating the project is better than proceeding further.

Third, even though the research process is a logical sequence of well-defined stages, it is not totally scientific. The subjective judgment of decision makers and researchers is likely to play a role in at least some of the stages, such as in the evaluation of the worth of a research project. Even mathematically sophisticated approaches for making cost-versus-benefit evaluations (such as the Bayesian analysis approach discussed in the appendix) are based on critical assumptions and require subjective inputs from decision makers. Despite the inevitability of some degree of subjectivity in the research process, marketing research should be conducted in a systematic fashion if it is to be beneficial.

The Wayne Engineering example illustrated the stages of the research process in a real-life setting. Of the several research features implicit in this example, especially noteworthy are the need for subjective inputs at the various stages, the interrelationships among the stages, the systematic way in which marketing research can be done, and the practical constraints researchers often face and must handle in some reasonable way.

KEY TERMS

Research process
Pseudo-Research
Secondary data
Primary data
Research proposal
Editing
Coding

QUESTIONS

1. "In conducting a marketing research project, researchers would be unwise to spend time and effort on subsequent stages unless all previous stages have been completed. Discuss this statement.
2. a. State, and briefly describe, the four criteria that are relevant in deciding whether or not a research project should be undertaken.
   b. One of the examples discussed in Chapter 1 dealt with Nissan Motor Company's research concerning relative awareness of the Datsun and Nissan names. Review this example, and state whether each of the four criteria was likely to have been met in this case. Justify your answer.
3. Explain the association between the nature of uncertainty faced by a decision maker and the potential benefit of a research project capable of reducing that uncertainty.

4. State whether you agree or disagree with this statement: "Correct identification and clear definition of the research problem are more crucial to the success of a project than sophisticated research techniques." Explain your answer.
5. Under what circumstances would a project be terminated without completing *all* the research process stages? Give an example of your own to illustrate your answer.
6. Does every marketing research project require a formal questionnaire? Discuss your answer, using one or two examples cited in previous chapters.
7. Considering all the facts given in the Wayne Engineering example, comment on the consultant's decision to restrict the study to cities and towns with populations of 50,000 or more. Specifically, what are the pros and cons of this decision?

[illegible] stages shown in Exhibit [illegible] through secondary sources [illegible] to design a sample, or collect primary data [illegible], when no trustworthy data [illegible] adequately meet a research [illegible] the project is better than proceeding [illegible] a logical sequence of well- [illegible]. The subjective judgment of decision makers and researchers also play a role in at least some [illegible] worth of a research project. [illegible] approaches for making cost-versus- [illegible] analysis approach discussed in the appendix [illegible] assumptions and require subjective inputs from decision makers. Despite the inevitability of some degree of subjectivity in the research process, marketing research should be conducted in a systematic fashion if it is to be beneficial.

The Wayne Engineering example illustrated the stages of the research process in a real-life setting. Of the several research features implicit in this example, especially noteworthy are the need for subjective inputs at the various stages, the interrelationships among the stages, the systematic way in which marketing research can be done, and the practical constraints researchers often face and must handle in some reasonable way.

KEY TERMS

Research process	Primary data	Editing
Pseudo-Research	Research proposal	Coding
Secondary data		

QUESTIONS

1. "In conducting a marketing research project, researchers would be unwise to spend time and effort on subsequent stages unless all previous stages have been completed." Discuss this statement.
2. State, and briefly describe, the four criteria that are relevant in deciding whether or not a research project should be undertaken.
3. One of the examples discussed in Chapter 1 dealt with Nissan Motor Company's research concerning relative awareness of the Datsun and Nissan names. Review this example, and state whether each of the four criteria was likely to have been met in this case. Justify your answer.
4. Explain the association between the nature of uncertainty faced by a decision maker and the potential benefit of a research project capable of reducing that uncertainty.

APPENDIX TO CHAPTER 3

Bayesian Analysis

This appendix describes a mathematical procedure for estimating the monetary value of a marketing research project. The procedure is commonly referred to as ***Bayesian analysis*** or the ***decision theory approach.***[1] We will use the situation depicted in scenario B presented earlier in the chapter to illustrate this procedure. In scenario B Jerry Wilson, advertising manager for corporation B, is wondering which of two commercials will be more appropriate for promoting his company's deodorant: commercial X, which uses a fear appeal, or commercial Y, which uses a romantic appeal.

Bayesian Analysis (Decision Theory Approach) *A mathematical procedure for estimating the monetary value of a marketing research project.*

WORTH OF PERFECT RESEARCH

Market States

The relative effectiveness of X or Y will hinge on the nature of the market; in particular, factors such as the current attitudes and the personalities of the target customers may determine which of the two commercials is likely to be more effective. For instance, commercial X (fear appeal) may be more effective if most of the target customers have low self-confidence, a personality trait. But if the target customers have high self-confidence, the fear appeal may backfire, while the romantic appeal may be effective. Thus the relative effectiveness of each commercial will depend on the overall nature or disposition of the market, or ***market state.*** For the sake of simplicity let us assume the market state in scenario B will be either M_1 or M_2, where M_1 and M_2 are defined as follows:

Market State *The overall nature or disposition of the market.*

---

[1] We will be discussing only the very basic elements of the decision theory approach in this appendix. Further details and refinements can be found in several books on the subject, including Joseph W. Newman, *Management Applications of Decision Theory* (New York: Harper & Row, 1971); Robert Schlaifer, *Analysis of Decisions Under Uncertainty* (New York: McGraw-Hill, 1969).

- M_1 is the state in which customer reaction will be favorable to commercial X but not to commercial Y, that is, a state in which the inherent persuasiveness of X is much higher than that of Y.
- M_2 is the state in which customer reaction will be favorable to commercial Y but not to commercial X, that is, a state in which the inherent persuasiveness of Y is much higher than that of X.

Prior Probabilities

As we know, Mr. Wilson is *uncertain* about which of the two market states, M_1 or M_2, actually exists. However, suppose Mr. Wilson's intuition tells him M_1 is slightly more likely to exist than M_2. Specifically, suppose he feels there is a 60% chance that M_1 is the true market state and a 40% chance that M_2 is the true market state. On the basis of these subjective feelings, we can specify the probabilities P of the two possible market states as

$$P(M_1) = .6,$$
$$P(M_2) = .4.$$

In decision theory these probabilities are called ***prior probabilities.***

Prior Probability *The probability that a particular market state or condition exists, based purely on the subjective feelings of a decision maker.*

Payoff Table *A table summarizing the net profits (or losses) under various possible combinations of decision alternatives and market states.*

Payoff Table

Mr. Wilson believes commercial X will result in a net profit to corporation B if market state M_1 exists but will lead to a net loss if M_2 exists. Similarly, he feels commercial Y will be beneficial only if market state M_2 exists. Mr. Wilson's quantified estimates of the net profits (or losses) under the various possible circumstances are summarized in Table A.3.1. A table like A.3.1 is called a ***payoff table.*** It indicates the payoffs, or monetary returns, associated with each alternative course of action A (A_1 or A_2 in our example) under each possible market state.

Expected Value Under Uncertainty

Choosing between alternatives A_1 and A_2 *without* conducting marketing research will amount to decision making *under uncertainty.* If Mr. Wilson wants to make his decision under uncertainty, what criterion should he use to choose between A_1 and A_2? The criterion commonly employed under such circumstances is the *expected value* of the alternative, which is its *weighted average payoff.* The expected value EV of A_1 and A_2 can be computed as follows:

$$\begin{aligned}\mathrm{EV}(A_1) &= (\text{payoff for } A_1 \textit{ if } M_1 \text{ exists}) \times P(M_1)\\ &\quad + (\text{payoff for } A_1 \textit{ if } M_2 \text{ exists}) \times P(M_2)\\ &= (800{,}000) \times (.6) + (-20{,}000) \times (.4)\\ &= 480{,}000 - 8000 = \$472{,}000,\\ \mathrm{EV}(A_2) &= (\text{payoff for } A_2 \textit{ if } M_1 \text{ exists}) \times P(M_1)\\ &\quad + (\text{payoff for } A_2 \textit{ if } M_2 \text{ exists}) \times P(M_2)\\ &= (-50{,}000) \times (.6) + (900{,}000) \times (.4)\\ &= -30{,}000 + 360{,}000 = \$330{,}000.\end{aligned}$$

As the computations indicate, the expected value of an alternative takes into account not only the payoff associated with it but also the probability of that payoff materializing. Thus an expected value can be interpreted as the overall worth of an alternative. Another way of interpreting the expected value of an

Table A.3.1 *Payoff Table in Scenario B*

	Payoffs Under Market State	
Alternative Courses of Action	M_1	M_2
Use commercial X (alternative A_1)	\$800,000	−\$20,000
Use commercial Y (alternative A_2)	−\$50,000	\$900,000

alternative is as follows: If one selects this same alternative under each of a large number of similar decision-making circumstances over a period of time, the long-run average payoff per decision will be equal to the expected value. Intuitively, the best alternative under a given set of circumstances is one with the highest expected value. Let us define ***expected value under uncertainty*** as the *maximum* expected payoff to a decision maker who chooses a course of action under uncertainty. We will use the notation EV(UC) to denote expected value under uncertainty.

Expected Value Under Uncertainty *The maximum expected payoff to a decision maker who chooses a course of action under uncertainty.*

In scenario B the expected value under uncertainty is given by

$$\begin{aligned} \text{EV(UC)} &= \text{maximum } [\text{EV}(A_1), \text{EV}(A_2)] \\ &= \text{EV}(A_1) = \$472{,}000. \end{aligned}$$

In other words, A_1 is the best alternative Mr. Wilson can choose under uncertainty, and his expected return is \$472,000.

Expected Value Under Certainty

Suppose David Cox, a marketing researcher, offers to conduct some *perfect* research to completely eliminate Mr. Wilson's uncertainty. Specifically, Mr. Cox tells Mr. Wilson: "My research will indicate to you for sure whether M_1 or M_2 is the true market state. In other words, if my research indicates customer reaction to be favorable to commercial X (i.e., market state M_1 exists), you can use commercial X and be *certain* your return will be \$800,000. But if my research indicates customer reaction to be unfavorable to commercial X (i.e., market state M_2 exists), you can employ commercial Y and be certain about obtaining a return of \$900,000." If you were Mr. Wilson, how much would you be willing to pay for Mr. Cox's research?

To answer this question, you must estimate the probability of the potential research outcome, indicating the market state to be M_1 or M_2. In other words, you must estimate the research outcome probabilities, $P(R_1)$ and $P(R_2)$, defined as follows:

$$\begin{aligned} P(R_1) &= P(\text{research outcome showing } M_1 \text{ is the market state}), \\ P(R_2) &= P(\text{research outcome showing } M_2 \text{ is the market state}). \end{aligned}$$

How do we estimate $P(R_1)$ and $P(R_2)$? Bear in mind that Mr. Cox *has not yet* conducted his research; hence at this point you do not know which research outcome (R_1 or R_2) will materialize. You only know that if Mr. Cox's perfect research, when completed, indicates a market state (M_1 or M_2), you can be certain that the indicated state is the true market state. Consequently, your best estimate of the probability of the research outcome indicating a particular mar-

Table A.3.2 *Decision Making Under Certainty*

Perfect Research Outcomes	True Market State	Best Course of Action	Monetary Return	P (Obtaining Return)
R_1	M_1	A_1	\$800,000	$P(R_1) = .6$
R_2	M_2	A_2	\$900,000	$P(R_2) = .4$

ket state is merely your *prior probability* estimate for that market state. That is,

$$P(R_1) = P(M_1) = .6,$$
$$P(R_2) = P(M_2) = .4.$$

Which commercial (X or Y) should Mr. Wilson use *if* the outcome of Mr. Cox's perfect research is R_1? Obviously, Mr. Wilson should use X. If R_1 occurs, Mr. Wilson can be sure M_1 exists, because of the perfect nature of the research, and hence using X can guarantee a return of \$800,000. Similarly, if R_2 occurs, using Y can guarantee a return of \$900,000.

Mr. Wilson will be making his decision *under certainty* if he asks Mr. Cox to conduct the perfect research, observes the research outcome, and then chooses the appropriate course of action. Table A.3.2 summarizes the two possible decision sequences and the corresponding returns when Mr. Wilson makes his decision under certainty.

Expected Value Under Certainty *The expected payoff to a decision maker who commissions a perfect research project, observes the research outcome, and chooses the best course of action corresponding to that outcome.*

Using the information in Table A.3.2, we can compute the ***expected value under certainty,*** EV(C), as follows:

$$\begin{aligned}\text{EV(C)} &= (\text{ultimate payoff if } R_1 \text{ occurs}) \times P(R_1)\\ &\quad + (\text{ultimate payoff if } R_2 \text{ occurs}) \times P(R_2)\\ &= (800{,}000) \times (.6) + (900{,}000) \times (.4)\\ &= 480{,}000 + 360{,}000 = \$840{,}000.\end{aligned}$$

The EV(C) can be interpreted as the expected return to a decision maker who commissions a perfect research project, observes the research outcome, and chooses the best course of action corresponding to that outcome.

Expected Value of Perfect Information

Expected Value of Perfect Information *The value obtained by subtracting the expected value under uncertainty from the expected value under certainty; this value represents the maximum amount a decision maker should be willing to pay for any research aimed at reducing uncertainty in a situation.*

Let us now turn our attention to estimating how much Mr. Wilson should be willing to pay for Mr. Cox's perfect research. Our discussion thus far indicates two possible ways for Mr. Wilson to make his decision: *under uncertainty,* by choosing between commercials X and Y without doing any research, or *under certainty,* by observing the outcome of perfect research first and then choosing between X and Y. The worth of perfect research should logically be reflected by how much better off Mr. Wilson would be by making his decision under certainty rather than under uncertainty. We already know the best returns to Mr. Wilson under these two options, namely, EV(C) and EV(UC), respectively. Hence the worth of perfect research, commonly known as the ***expected value of perfect information,*** EV(PI), is given by

$$\text{EV(PI)} = \text{EV(C)} - \text{EV(UC)} = 840{,}000 - 472{,}000 = \$368{,}000.$$

The EV(PI) can be interpreted as the *maximum* amount a decision maker should be willing to pay for *any* research promising to reduce his or her uncertainty. In other words, EV(PI) provides an upper bound for the worth of marketing research in a given situation. No real-life marketing research is likely to be perfect; there is always room for error. In our example Mr. Cox's promise to conduct perfect research is more hypothetical than real. Consequently, if Mr. Wilson were to commission a real-life research project to aid his decision making, he should pay no more than $368,000; indeed, realistically speaking, he should only pay some amount *less than* the upper limit of $368,000. How much less should he pay? The answer depends on how "imperfect" the proposed research is likely to be, or how much room exists for error. We will now discuss a procedure for determining the worth of research that is less than perfect.

WORTH OF IMPERFECT RESEARCH

Conditional Probabilities of Research Outcomes

Suppose Mr. Wilson does not believe Mr. Cox's claim that his research will be perfect. Specifically, on the basis of past experience with research projects of the type proposed by Mr. Cox, Mr. Wilson believes there is only a 70% chance that Mr. Cox's research will identify the existence of M_1 when the true market state is M_1. In other words, when the true market state is that consumer reaction will be favorable to commercial X, there is only a .7 probability that Mr. Cox's research outcome will be R_1; alternatively, there is a .3 probability that the research outcome will be R_2. These probabilities are ***conditional probabilities,*** which can be represented as follows:

Conditional Probability
The probability that a research outcome would be observed if a particular market state or condition were to exist.

$$P(R_1/M_1) = P(\text{obtaining research outcome } R_1 \textit{ if} \text{ the market state is } M_1) = .7,$$
$$P(R_2/M_1) = P(\text{obtaining research outcome } R_2 \textit{ if} \text{ the market state is } M_1) = .3.$$

Similarly, suppose Mr. Wilson feels there is only an 80% chance that the research will correctly identify the market state when the true market state is M_2. Then the conditional probabilities of the research outcomes are

$$P(R_1/M_2) = .2 \quad \text{and} \quad P(R_2/M_2) = .8.$$

Posterior Probabilities

We know that the prior probability of market state M_1, $P(M_1)$, is .6, on the basis of Mr. Wilson's subjective feelings. Suppose the imperfect research described in the previous section is conducted and outcome R_1 is observed. Would this research information (i.e., outcome R_1) have an impact on Mr. Wilson's assessment of the prior probabilities of market states M_1 and M_2? Yes. Since outcome R_1 has been observed, we would intuitively expect Mr. Wilson to *increase* his estimate of the probability that M_1 is the true market state and *decrease* his estimate of the probability that M_2 is the true market state.

Posterior Probability
The probability for a market state or condition that is contingent upon a particular research outcome being observed.

The *revised* probabilities of the market states, upon observation of a particular research outcome, are called ***posterior probabilities.*** Note that posterior probabilities are contingent on a particular research outcome. Hence each possible research outcome will have a corresponding set of posterior probabilities of the market states. In our example there are two sets of posterior probabilities: (1) the posterior probabilities of M_1 and M_2 when the research outcome is R_1, denoted as $P(M_1/R_1)$ and $P(M_2/R_1)$; and (2) the posterior probabilities of M_1 and M_2 when the research outcome is R_2, denoted as $P(M_1/R_2)$ and $P(M_2/R_2)$.

By what amount should Mr. Wilson revise the prior probabilities in order to obtain the posterior probabilities? The posterior probabilities can be derived from the prior and conditional probabilities by using a principle known as *Bayes's rule.* According to Bayes's rule, for any research outcome R_i and market state M_j,

$$P(M_j/R_i) = \frac{P(R_i/M_j) \times P(M_j)}{P(R_i)} \tag{A.3.1}$$

In Eq. (A.3.1), $P(M_j/R_i)$ is the posterior probability of market state M_j, given that research outcome R_i is observed; $P(R_i/M_j)$ is the conditional probability of R_i if the market state is M_j; $P(M_j)$ is the prior probability of market state M_j; and $P(R_i)$ is the probability that R_i will be the research outcome. If there are K possible market states $(M_1, M_2, \ldots, M_K)$, the probability of research outcome R_i is given by

$$\begin{aligned} P(R_i) = {} & P(\text{research outcome is } R_i \textit{ and } \text{the market state is } M_1) \\ & + P(\text{research outcome is } R_i \textit{ and } \text{the market state is } M_2) + \ldots \\ & + P(\text{research outcome is } R_i \textit{ and } \text{the market state is } M_K). \end{aligned}$$

Each term in this expression is a *joint probability* of research outcome R_i and a particular market state. Thus

$$P(R_i) = \sum_{j=1}^{K} P(R_i \textit{ and } M_j) = \sum_{j=1}^{K} P(R_i/M_j) \times P(M_j).$$

We can rewrite Eq. (A.3.1) as

$$P(M_j/R_i) = \frac{P(R_i/M_j) \times P(M_j)}{\sum_{j=1}^{K} P(R_i/M_j) \times P(M_j)} \tag{A.3.2}$$

Equation (A.3.2) can be used to calculate the various posterior probabilities.

Table A.3.3 shows the computation of the posterior probabilities of M_1 and M_2 under each of the two possible research outcomes (R_1 and R_2) in our example.

Decision Making Based on Imperfect Research

The posterior probabilities computed in Table A.3.3 will be helpful in determining the best alternative course of action (A_1 or A_2) under each of the two possible research outcomes. The criterion for choosing between A_1 and A_2 is the *expected value* of each, *given that a particular research outcome has been observed.* Such contingent expected values can be computed by using a procedure similar to the one presented in the section "Expected Value Under Uncertainty," *with one major difference*: We should use the *posterior* (rather than *prior*) probabilities of the market states in our computation. Table A.3.4 shows how the contingent expected values of A_1 and A_2 are calculated.

Expected Value of Decision Making with Imperfect Information *The expected payoff to a decision maker who commissions a less-than-perfect research project, observes the research outcome, and chooses the best course of action corresponding to that outcome.*

From the expected values in Table A.3.4 Mr. Wilson should choose alternative A_1 *if* the research outcome is R_1; he should choose alternative A_2 *if* the research outcome is R_2. The corresponding expected returns are \$668,800 and \$558,000, respectively. But bear in mind that the research project has not yet been conducted; hence we do not know which outcome will emerge if the research is conducted. However, we do know the probabilities of each possible research outcome occurring (see Table A.3.3). Therefore we can compute the ***expected value of decision making with imperfect information*** as follows:

Table A.3.3 *Computation of Posterior Probabilities*

Research Outcomes	Prior Probabilities	Conditional Probabilities	Joint Probabilities and Research Outcome Probabilities	Posterior Probabilities
R_1	$P(M_1) = .6$	$P(R_1/M_1) = .7$	$P(R_1 \text{ and } M_1) = P(R_1/M_1) \times P(M_1)$ $= (.7) \times (.6)$ $= .42$	$P(M_1/R_1) = P(R_1 \text{ and } M_1)/P(R_1)$ $= .42/.50$ $= .84$
	$P(M_2) = .4$	$P(R_1/M_2) = .2$	$P(R_1 \text{ and } M_2) = P(R_1/M_2) \times P(M_2)$ $= (.2) \times (.4)$ $= .08$	$P(M_2/R_1) = P(R_1 \text{ and } M_2)/P(R_1)$ $= .08/.50$ $= .16$
			$P(R_1) = .42 + .08 = .50$	
R_2	$P(M_1) = .6$	$P(R_2/M_1) = .3$	$P(R_2 \text{ and } M_1) = P(R_2/M_1) \times P(M_1)$ $= (.3) \times (.6)$ $= .18$	$P(M_1/R_2) = P(R_2 \text{ and } M_1)/P(R_2)$ $= .18/.50$ $= .36$
	$P(M_2) = .4$	$P(R_2/M_2) = .8$	$P(R_2 \text{ and } M_2) = P(R_2/M_2) \times P(M_2)$ $= (.8) \times (.4)$ $= .32$	$P(M_2/R_2) = P(R_2 \text{ and } M_2)/P(R_2)$ $= .32/.50$ $= .64$
			$P(R_2) = .18 + .32 = .50$	

$$\begin{aligned}
\text{EV(DM with II)} &= \text{EV(decision making with imperfect information)} \\
&= (\text{ultimate payoff if } R_1 \text{ occurs}) \times P(R_1) \\
&\quad + (\text{ultimate payoff if } R_2 \text{ occurs}) \times P(R_2) \\
&= (668{,}800) \times (.5) + (558{,}000) \times (.5) \\
&= 334{,}400 + 279{,}000 = \$613{,}400.
\end{aligned}$$

The EV(DM with II) can be interpreted as the expected return to a decision maker who commissions a research project that is less than perfect, observes the research outcome, and chooses the best course of action corresponding to that outcome.

Expected Value of Imperfect Information

How much should Mr. Wilson be willing to pay for the less-than-perfect research project? The answer depends on how much better off Mr. Wilson would be by making a decision on the basis of the research information rather than under complete uncertainty. Specifically, the ***expected value of imperfect information*** is given by

Expected Value of Imperfect Information *The value obtained by subtracting the expected value under uncertainty from the expected value of decision making with imperfect information; this value represents the monetary worth of a proposed research project that is less than perfect.*

$$\begin{aligned}
\text{EV(II)} &= \text{EV(DM with II)} - \text{EV(UC)} \\
&= 613{,}400 - 472{,}000 = \$141{,}400.
\end{aligned}$$

The EV(II) is the monetary estimate of the worth of a research project to a decision maker. In our example, if the cost of the proposed research is less than \$141,400, for Mr. Wilson to undertake the research project will be worthwhile. If the cost is greater than \$141,400, Mr. Wilson will be better off choosing the best alternative *without* doing any research—namely, A_1, as we saw earlier.

Table A.3.4 *Contingent Expected Values*

Expected Values Contingent Upon Research Outcome R_1	Expected Values Contingent Upon Research Outcome R_2
$\text{EV}(A_1/R_1) = (\text{payoff for } A_1 \textit{ if } M_1 \text{ exists}) \times P(M_1/R_1)$ $+ (\text{payoff for } A_1 \textit{ if } M_2 \text{ exists}) \times P(M_2/R_1)$ $= (800{,}000) \times (.84) + (-20{,}000) \times (.16)$ $= 672{,}000 - 3200 = \$668{,}800$	$\text{EV}(A_1/R_2) = (\text{payoff for } A_1 \textit{ if } M_1 \text{ exists}) \times P(M_1/R_2)$ $+ (\text{payoff for } A_1 \textit{ if } M_2 \text{ exists})$ $\times P(M_2/R_2)$ $= (800{,}000) \times (.36) + (-20{,}000) \times (.64)$ $= 288{,}000 - 12{,}800 = \$275{,}200$
$\text{EV}(A_2/R_1) = (\text{payoff for } A_2 \textit{ if } M_1 \text{ exists}) \times P(M_1/R_1)$ $+ (\text{payoff for } A_2 \textit{ if } M_2 \text{ exists}) \times P(M_2/R_1)$ $= (-50{,}000) \times (.84) + (900{,}000) \times (.16)$ $= -42{,}000 + 144{,}000 = \$102{,}000$	$\text{EV}(A_2/R_2) = (\text{payoff for } A_2 \textit{ if } M_1 \text{ exists}) \times P(M_1/R_2)$ $+ (\text{payoff for } A_2 \textit{ if } M_2 \text{ exists})$ $\times P(M_2/R_2)$ $= (-50{,}000) \times (.36) + (900{,}000) \times (.64)$ $= -18{,}000 + 576{,}000 = \$558{,}000$

DECISION TREES

Decision Tree *A diagram that lays out, in logical sequence, various available decision options, intervening events, and/or market states that can influence the monetary consequences of choosing each option.*

A ***decision tree*** is a diagram that presents the following information in a logical sequence:

1. The decision options (or alternative courses of action) available to a decision maker.
2. The intervening events and/or market states that can influence the monetary consequences of choosing each option.

Exhibit A.3.1 is a decision tree summarizing the situation facing Mr. Wilson. Tracing the topmost branch of the tree in Exhibit A.3.1, for instance, we see that *if* Mr. Wilson decides not to conduct any marketing research and chooses alternative A_1, and *if* the market state is M_1, the monetary returns to Mr. Wilson will be $800,000.

Exhibit A.3.2 is a detailed version of Exhibit A.3.1. It is a decision tree showing the expected values of each decision option as well as the appropriate probabilities of the research outcomes and market states. The upper segment of the tree relates to decision making under uncertainty: the lower segment relates to decision making with imperfect information. At any node in the tree from which more than one decision branch emerges, the best branch to take is the one with the *highest expected value*. For instance, if Mr. Wilson commissions the research project and outcome R_2 is observed, he should choose alternative A_2.

Constructing a decision tree is not totally different from the computational procedure we discussed earlier for evaluating different decision options and estimating the worth of research information. Indeed, in order to draw a tree with all its branches labeled as shown in Exhibit A.3.2, we would have to go through all the necessary probability and expected value computations. The primary benefit of a decision tree diagram is that it provides a pictorial summary that enables a decision maker to see, at a glance, what the various decision options are and the monetary consequence of choosing each.

Exhibit A.3.1 *Mr. Wilson's Decision Tree*

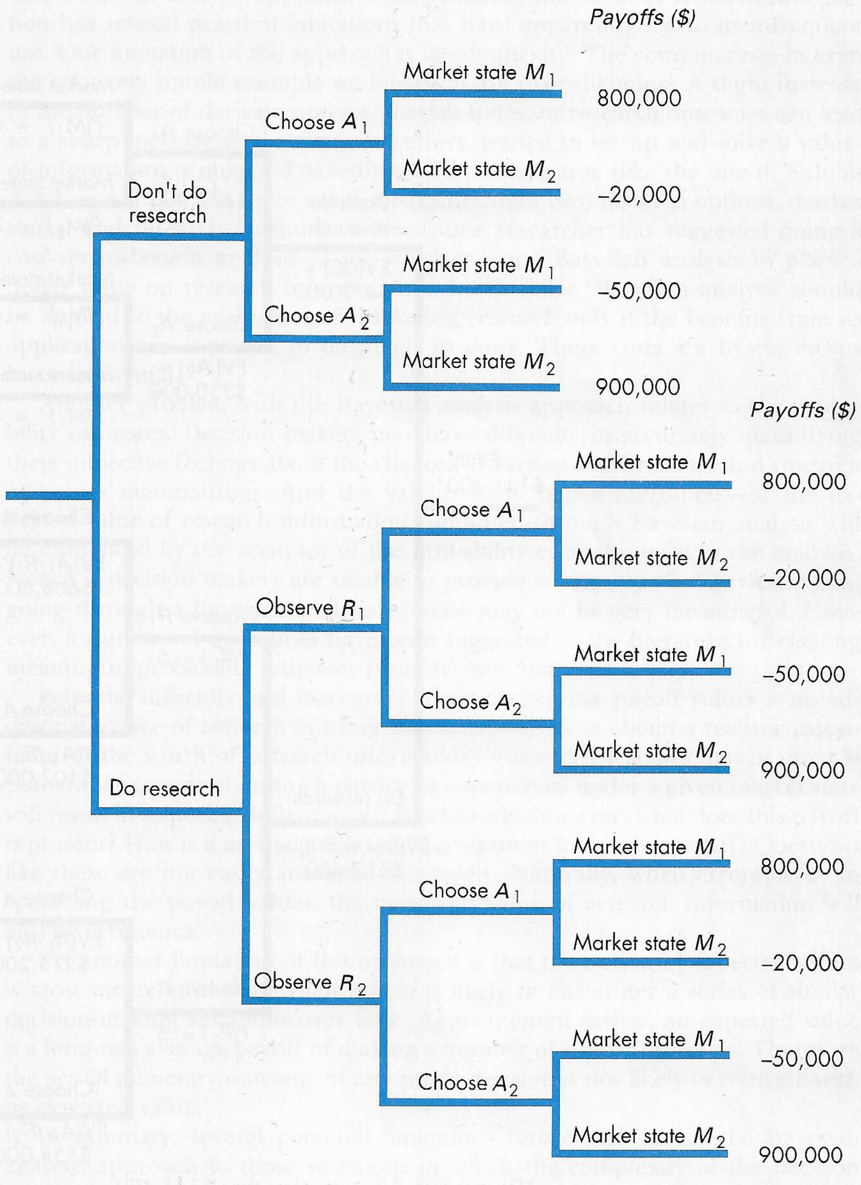
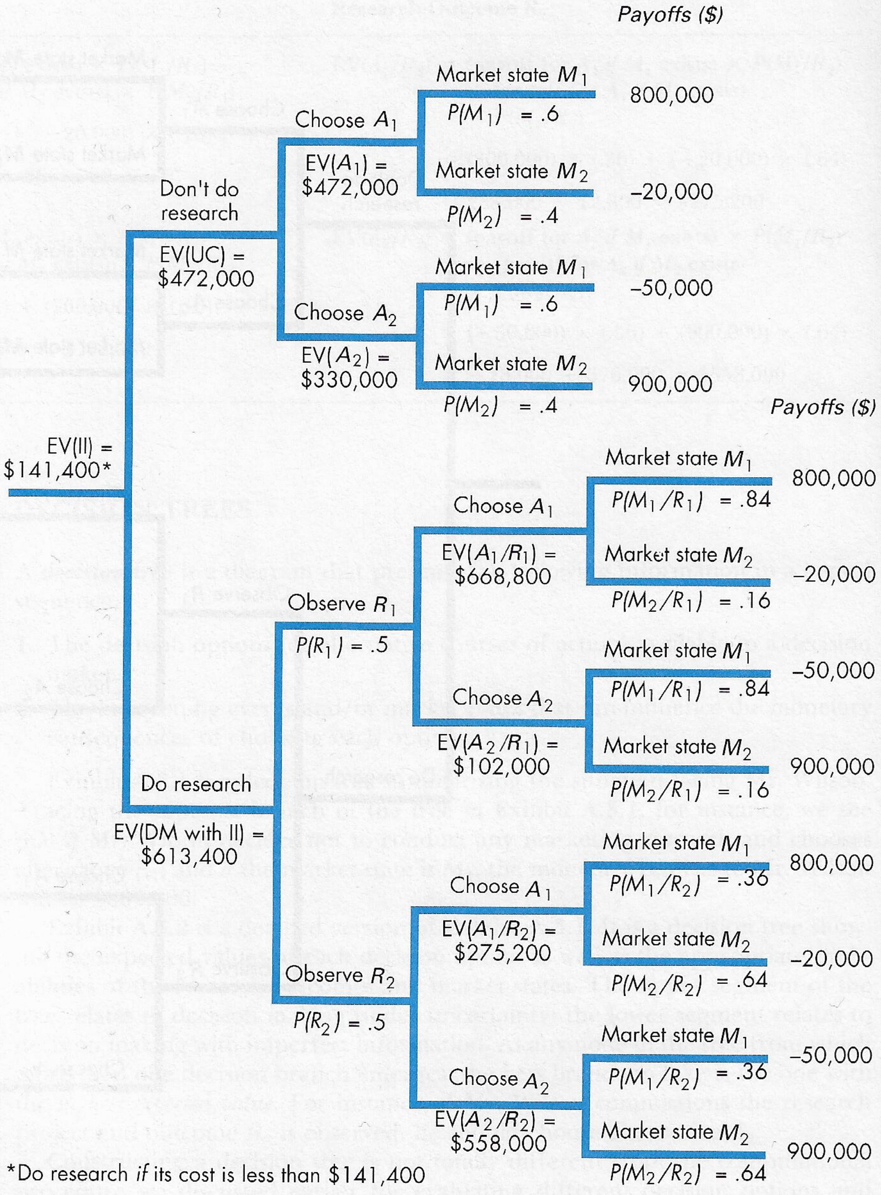

Payoffs ($)
Don't do research
Choose A_1
Market state M_1
800,000
Market state M_2
–20,000
Choose A_2
Market state M_1
–50,000
Market state M_2
900,000
Payoffs ($)
Do research
Observe R_1
Choose A_1
Market state M_1
800,000
Market state M_2
–20,000
Choose A_2
Market state M_1
–50,000
Market state M_2
900,000
Observe R_2
Choose A_1
Market state M_1
800,000
Market state M_2
–20,000
Choose A_2
Market state M_1
–50,000
Market state M_2
900,000

Exhibit A.3.2 *Decision Tree with Probabilities and Expected Values*


Payoffs ($)
Market state M_1
$P(M_1)$ = .6
800,000
Choose A_1
EV(A_1) = $472,000
Market state M_2
$P(M_2)$ = .4
–20,000
Don't do research
EV(UC) = $472,000
Market state M_1
$P(M_1)$ = .6
–50,000
Choose A_2
EV(A_2) = $330,000
Market state M_2
$P(M_2)$ = .4
900,000
Payoffs ($)
EV(II) = $141,400*
Market state M_1
$P(M_1/R_1)$ = .84
800,000
Choose A_1
EV(A_1/R_1) = $668,800
Market state M_2
$P(M_2/R_1)$ = .16
–20,000
Observe R_1
$P(R_1)$ = .5
Market state M_1
$P(M_1/R_1)$ = .84
–50,000
Choose A_2
EV(A_2/R_1) = $102,000
Market state M_2
$P(M_2/R_1)$ = .16
900,000
Do research
EV(DM with II) = $613,400
Market state M_1
$P(M_1/R_2)$ = .36
800,000
Choose A_1
EV(A_1/R_2) = $275,200
Market state M_2
$P(M_2/R_2)$ = .64
–20,000
Observe R_2
$P(R_2)$ = .5
Market state M_1
$P(M_1/R_2)$ = .36
–50,000
Choose A_2
EV(A_2/R_2) = $558,000
Market state M_2
$P(M_2/R_2)$ = .64
900,000

*Do research *if* its cost is less than $141,400

LIMITATIONS OF BAYESIAN ANALYSIS

The Bayesian analysis approach to determining the value of research information has several practical limitations that have apparently led to its infrequent use. One limitation of the approach is its complexity. The computations in even the relatively simple example we looked at are overwhelming. A slight increase in the number of decision options, market states, or research outcomes can lead to a sharp increase in the time and effort needed to set up and solve a value-of-information problem. Try setting up a decision tree (like the one in Exhibit A.3.1 or A.3.2) when there are three (rather than two) decision options, market states, and research outcomes. Indeed, one researcher has suggested doing a cost-versus-benefit analysis to see whether using Bayesian analysis to place a dollar value on research information is worthwhile: "Bayesian analysis should be applied to the evaluation of marketing research only if the benefits from its application are expected to outweigh its costs. These costs are by no means inconsequential."[2]

Another problem with the Bayesian analysis approach relates to the probability estimates. Decision makers may have difficulty in accurately quantifying their subjective feelings about the chances of various market states and research outcomes materializing. And the validity—and hence usefulness—of the expected value of research information computed through Bayesian analysis will be influenced by the accuracy of the probability estimates used in the analysis. Hence if decision makers are unable to provide accurate probability estimates, going through a Bayesian analysis exercise may not be very meaningful. However, a number of guidelines have been suggested in the literature for eliciting meaningful probability estimates from decision makers.[3]

Potential difficulty and inaccuracy in estimating the payoff values is an additional source of concern in using Bayesian analysis to obtain a realistic assessment of the worth of research information. Table A.3.1 shows that in our hypothetical example choosing a particular commercial under a given market state will result in a specific dollar payoff. Realistically, however, what does this payoff represent? How is it arrived at? Is it a short-term or long-term payoff? Questions like these are not easily answered accurately. Naturally, when errors occur in specifying the payoff values, the estimated value of research information will also be erroneous.

Yet another limitation of this approach is that the notion of expected values is most meaningful only when a firm is likely to encounter a series of similar decision-making situations over time. As mentioned earlier, an expected value is a long-run average payoff of making a number of similar decisions. Therefore the actual monetary outcome of any single decision is not likely to coincide with its expected value.

In summary, several potential limitations restrict the use of the Bayesian analysis approach to those situations in which the complexity of the decision problem is not overwhelming; the necessary probabilities and payoffs are easy to estimate with at least a reasonable degree of accuracy; and similar decision-making settings arise frequently. However, in practice, such situations are more an exception than the rule. Nevertheless, some aspects of the approach may still be helpful to decision makers in terms of organizing their thinking in a logical fashion. For instance, constructing a decision tree diagram (even when accurate quantitative estimates of probabilities and payoffs are not available) will force a decision maker to think through the various market states and possible outcomes methodically, prior to making a final choice. Such an exercise may prevent the decision maker from overlooking any important aspect of the decision problem.[4]

---

[2] Gert Assmus, "Bayesian Analysis for the Evaluation of Marketing Research Expenditures: A Reassessment," *Journal of Marketing Research,* 14 (November 1977), p. 566.

[3] See, for example, Rakesh Kumar Sarin, "Elicitation of Subjective Probabilities in the Context of Decision-Making," *Decision Sciences,* 9 (January 1978), pp. 37–48; Philip Kotler, "A Guide to Gathering Expert Estimates," *Business Horizons,* 13 (October 1970), pp. 79–87.

KEY TERMS

Bayesian analysis (decision theory approach)
Market state
Prior probability
Payoff table
Expected value under uncertainty
Expected value under certainty
Expected value of perfect information
Conditional probability
Posterior probability
Expected value of decision making with imperfect information
Expected value of imperfect information
Decision tree

QUESTIONS

1. In the context of the decision theory approach to evaluating the monetary worth of marketing research, clearly state, in a sentence or two, what each of the following terms means: (a) prior probability; (b) payoff table; (c) expected value under uncertainty; (d) expected value under certainty.
2. Since marketing research cannot realistically be perfect, what is the purpose in determining the expected value of perfect information? Discuss your answer.
3. In what ways do posterior probabilities differ from prior probabilities?
4. Briefly discuss the practical limitations of using Bayesian analysis to estimate the value of research information.
5. A marketer has two alternative promotion strategies, A_1 and A_2. The relative success of A_1 and A_2 depends on the attitudinal structure of the target market. From experience the marketer feels that the target market has a 60% chance of having attitudinal structure S_1 and a 40% chance of having attitudinal structure S_2 (assume that S_1 and S_2 are the only possible states of nature). The *incremental returns* from each alternative, depending on each state of nature, are given in the following payoff table:

	S_1	S_2
A_1	\$100,000	−\$50,000
A_2	−\$70,000	\$200,000

The marketer also has the option of doing a research project to reduce her uncertainty about the true states of nature. The research project has two possible outcomes, Z_1 and Z_2. From past experience the potential accuracy of the research results are given in terms of the following conditional probabilities:

---

[4] Paul E. Green offers a concise discussion of the limitations and potential strengths of the Bayesian analysis approach in the concluding section of his article "Bayesian Decision Theory in Pricing Strategy," *Journal of Marketing*, 27 (January 1963), pp. 5–14.

$$P(Z_1/S_1) = .8, \qquad P(Z_1/S_2) = .2,$$
$$P(Z_2/S_1) = .2, \qquad P(Z_2/S_2) = .8.$$

a. Determine EV(PI) and EV(DM with II).
b. Construct a decision tree for this problem, and represent the appropriate probabilities, expected values, and so on.

6. You are given the following information: The three possible states of nature are S_1, S_2, and S_3. The prior probabilities are $P(S_1) = .4$, $P(S_2) = .4$, and $P(S_3) = .2$. The three decision alternatives are A_1, A_2, and A_3. The three possible outcomes if marketing research is done are Z_1, Z_2, and Z_3. The payoff table (all numbers are in dollars) is as follows:

	S_1	S_2	S_3
A_1	−50,000	0	200,000
A_2	0	200,000	−50,000
A_3	100,000	50,000	0

The conditional probability table is as follows:

	S_1	S_2	S_3
Z_1	.6	.2	.1
Z_2	.3	.6	.1
Z_3	.1	.2	.8
	1.0	1.0	1.0

Determine the following information:
a. The best alternative to choose with no research.
b. The EV(PI).
c. The EV(II).
d. The best alternative to choose if the firm cannot afford to lose any money.

7. You are given the accompanying payoff table related to four states of nature and three alternative courses of action. Each state of nature is equally likely; that is, the prior probability for each state of nature is .25. What is the EV(PI), given this payoff table?

	S_1	S_2	S_3	S_4
A_1	50	50	40	80
A_2	60	70	50	100
A_3	30	60	30	90

(*Hint:* This problem can be solved merely by inspection.)

8. A cosmetics manufacturing firm that introduces several new products a year is wondering whether or not to introduce a new hair-coloring product it has developed. From past experience with similar products, management feels that the new product has a 70% chance of success and a 30% chance of failure. If the product is introduced full-scale and it succeeds, the firm stands to make a net profit of $500,000. If the product fails, however, the firm will lose $300,000. Before full-scale introduction the firm has the option of test-marketing the product at a cost of about $85,000. Should it test-market the product? What other recommendation would you make concerning this new product? Explain your answer.

CHAPTER 4

Types of Marketing Research

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between exploratory research and conclusive research and discuss the role played by each in research projects.*
- *Illustrate four approaches for conducting exploratory research.*
- *Discuss the nature of the distinction between descriptive research and experimental research.*
- *Distinguish between a cross-sectional study and a longitudinal study and state when each type of study is appropriate.*
- *Distinguish between (a) a cross-sectional sample and a panel, and (b) an omnibus panel and a true panel.*
- *State the advantages and drawbacks of a longitudinal study that uses a true panel versus one that uses a series of cross-sectional samples.*

Chapter 3 described the general sequence of steps followed in conducting any marketing research project. However, the nature and type of marketing research varies depending on the unique characteristics of a given situation. Consider the following situations and try to identify similarities and differences among them.

Situation A. Modern Office Products, Inc. (MOP), manufactures a broad line of office equipment and supplies. It sells its products to a variety of organizations through its sales force. Despite a healthy growth in industry sales, MOP's sales and profits have declined during the past two years, much to the concern of MOP executives.

Situation B. Saver's National Bank (SNB) has grown rapidly since its inception a few years ago. Its growth is apparently due to a unique set of financial services it offers. While SNB management is pleased with the bank's performance thus far, it is worried about growing competition from a variety of financial institutions. To consolidate SNB's current market position, the bank's executives want to ascertain the demographic composition of customers and their perceptions about the bank's strengths and weaknesses.

Situation C. Trent Eating Association (TEA) operates a chain of restaurants in eight communities of similar size and population characteristics. The TEA currently has the image of a high-class restaurant chain serving excellent food at premium prices. Its president is wondering whether a 15% reduction in prices of all menu items would hurt, or help, sales revenues and profits.

EXPLORATORY VERSUS CONCLUSIVE RESEARCH

A common feature across situations A, B, and C is the need for marketing research. The managements of MOP, SNB, and TEA are faced with uncertainties capable of being reduced through marketing research. However, if you examine the three situations carefully, you will also notice some differences. For instance, the purpose of any potential marketing research is less clear in the situation facing MOP than in the situations facing SNB and TEA.

The goal of the marketing research conducted in situation A should be to aid MOP's management in arresting the decline in sales and profits. However, some preliminary investigation should be done to identify the right problem to be researched. Could MOP's poor market performance be a result of deteriorating quality of its products? Could it be due to lack of adequate motivation on the part of MOP's sales force? Could it be due to ineffective promotion and the resultant weak customer awareness of and loyalty toward MOP and its products? The MOP executives must examine such questions to pinpoint the most fruitful avenue for further research. Preliminary investigation of this nature is conventionally labeled exploratory research.

Definition. ***Exploratory research*** is research intended to develop initial hunches or insights and to provide direction for any further research needed.

alku

aavistus =

The primary purpose of exploratory research is to shed light on the nature of a situation and identify any specific objectives or data needs to be addressed through additional research. Exploratory research is most useful when a decision maker wishes to better understand a situation and/or identify decision alternatives.

Unlike the situation being faced by MOP, the SNB and TEA scenarios imply a much clearer definition of research purpose and data requirements. The bank is particularly interested in obtaining a demographic profile of its customer base and uncovering customer perceptions of SNB's strengths and weaknesses. Research in this case should therefore focus on obtaining and analyzing demographic and perceptual data from a cross section of SNB's customers. Such research should, in turn, aid SNB in formulating a strategy to consolidate its market position. Trent's president is interested in determining the impact of a price reduction on sales revenues and profits. The situation here is even more clear-cut, and a decision about whether or not to reduce prices can be made after marketing research determines the effect of a price reduction. Therefore in both the SNB and TEA scenarios data requirements are clearer than in the MOP scenario, and the findings are more likely to lead to final decisions. Research of this nature is usually termed conclusive research.

Table 4.1 *Differences Between Exploratory and Conclusive Research*

Research Project Components	Exploratory Research	Conclusive Research
Research purpose ymmärrys, käsitys =	General: To generate insights about a situation	Specific: To verify insights and aid in selecting a course of action
Data needs	Vague	Clear
Data sources	Ill-defined	Well-defined
Data collection form	Open-ended, rough	Usually structured
Sample	Relatively small; subjectively selected to maximize generation of useful insights	Relatively large; objectively selected to permit generalization of findings
Data collection	Flexible; no set procedure	Rigid; well-laid-out procedure
Data analysis	Informal; typically non-quantitative	Formal; typically quantitative
Inferences/ recommendations	More tentative than final	More final than tentative

Definition. *Conclusive research* is research intended to verify insights and to aid decision makers in selecting a specific course of action.

The primary purpose of conclusive research, sometimes also known as *confirmatory research*, is to help decision makers choose the best course of action in a situation. It is especially useful when a decision maker already has in mind one or more alternatives and is specifically looking for information pertinent to evaluating them. Since data requirements are clearly specified in a situation calling for conclusive research, and since such research is intended as an aid in final stages of the decision-making process, it is typically more formal and rigorous than exploratory research.

In short, the purpose of a research project and the preciseness of its data requirements determine whether the research is exploratory or conclusive. Exploratory research is conducted when decision makers sense a need for marketing research but are unsure of the specific direction the research should take. Conclusive or confirmatory research is conducted when decision makers have a clearer idea about the types of information they want.

Even though exploratory and conclusive research have distinct purposes, they both consist of the same research components. They differ only in terms of the degree of formalization and flexibility of the components, as illustrated by Table 4.1. Exploratory-research projects are quite informal and flexible, as discussed next.

CONDUCTING EXPLORATORY RESEARCH

Interviews with Knowledgeable Individuals

An effective way of doing exploratory research is to seek out and talk to individuals with expertise in areas related to the situation being investigated. For instance, in situation A described earlier, MOP executives can gain useful insights by informally discussing MOP's deteriorating market performance with a few key salespeople and customers. Discussions with knowledgeable salespeople, for example, may reveal whether lack of sales force motivation is a problem and, if so, whether it is serious enough to warrant further investigation. Conducting exploratory research by interviewing knowledgeable individuals is sometimes called the ***key-informant technique***[1]. Some books also refer to it as an *expert-opinion survey*.

Key-Informant Technique (Expert-Opinion Survey) *Conducting exploratory research by seeking out and talking to individuals with expertise in areas related to the situation being investigated.*

The key-informant technique is necessarily a very subjective and flexible procedure with no standard approach. However, proper selection of informants is crucial to obtaining useful information with a minimum of time and effort. In discussing the use of the key-informant technique in industrial-marketing settings, William Cox states: "It is not uncommon to find that less than 1% of the 'knowledgeable' persons associated with an industrial market possess virtually all of the relevant information about the market. . . [Hence] careful attention [must] be given to the selection of knowledgeable persons, using judgment samples rather than any random sampling method."[2] Indeed, Cox's observation and advice are not limited to industrial marketing settings; they are relevant in almost every context requiring exploratory research.

The key-informant technique, if properly used, can be very productive in situations where a decision maker senses the need for research but does not have well-defined research objectives. Table 4.2 presents several such situations, along with examples of knowledgeable individuals capable of providing valuable insights in each situation.

---

[1] Michael J. Houston, "The Key Informant Technique: Marketing Application," in *Conceptual and Methodological Foundations of Marketing*, ed. Thomas V. Greer (Chicago: American Marketing Association, 1974), pp. 305–308, contains a good discussion of the use of this technique in marketing contexts.

[2] William E. Cox, Jr., *Industrial Marketing Research* (New York: Wiley, 1979), pp. 22–23. Claire Selltiz, Lawrence S. Wrightsman, and Stuart W. Cook, *Research Methods in Social Relations* (New York: Holt, Rinehart and Winston, 1976), also emphasizes the importance of carefully selecting key informants.

Table 4.2 *Examples of Knowledgeable Individuals*

Situation Calling for Exploratory Research	Knowledgeable Individuals with Useful Insights
A company's brand X detergent, a leader in the detergent market, is suddenly losing market share	Key managers in wholesale and retail firms through which the company distributes brand X
A newly formed nonprofit organization to aid people with serious physical handicaps is wondering about the charitable programs it should develop and the strategies it should use to seek donations from the public	Officers in well-established public service organizations, such as the United Way, the American Red Cross, and the Muscular Dystrophy Association
A farm equipment manufacturer is thinking about developing a new line of completely automatic, programmable, seed planters	Farm equipment dealers, farmers with huge farms, and farmers currently using sophisticated farm equipment
A United States presidential candidate is wondering how to allocate campaign resources efficiently across various regions of the country	State and local party officials with knowledge about public sentiment and attitudes in their respective areas

Focus Group Interviews

Focus Group Interview
A setting in which respondents (typically about 8 to 12) discuss a given topic in a fairly informal fashion in the presence of a well-trained, objective moderator.

Talking to experts or knowledgeable individuals is not the only way of gaining a better understanding of a situation calling for research. Another frequently used informal-interviewing method is the ***focus group interview***, which is sometimes simply called focus group. In a focus group interview respondents (typically about 8 to 12) discuss a given topic in a fairly informal fashion. The discussion is led by a well-trained researcher, called a moderator. The moderator's primary tasks are to ensure that key aspects of the topic get discussed and to observe/record the reactions of the participants. Focus groups are used in a variety of situations, such as for gaining insights about consumer acceptance of a new product idea, about potential reactions to an advertising theme, or about types of questions to include in a consumer survey and how to word them. A major portion of Chapter 7 is devoted to focus group interviews.

Analysis of Secondary Data

Secondary Data
Data collected for some purpose other than the research situation.

Secondary data, as we saw in the previous chapter, are data that have already been collected. Examining appropriate secondary data is a fast and inexpensive way of doing exploratory research. Analysis of secondary data relating to situations requiring marketing research can generate valuable insights. Such insights, in turn, will provide a proper focus for conclusive research. Sometimes, the insights revealed by secondary-data

Research in Use 4.1 *Blue Cross and Blue Shield of Iowa*

Use of Secondary-Data Analysis in Decision Making

In 1975 Blue Cross and Blue Shield of Iowa introduced a new program, the School Delta Dental Accident Program, offering dental-injury coverage to students, faculties, and staffs of schools. The program attracted an initial enrollment of 73 schools with 11,143 participants and quickly grew to a peak enrollment of 262 schools with 20,457 participants. However, by the end of the 1981–1982 school year, enrollment in the program had dropped to 150 schools with only 5638 participants.

The marketing research division of Blue Cross and Blue Shield analyzed a variety of secondary data in order to determine possible reasons for the decline in enrollment. Several enlightening revelations emerged from this analysis:

- Because of increases in taxes and inflation, the average American worker experienced a decline in real pay in the late 1970s. This observation emerged from data published in the *U.S. News and World Report* (November 9, 1981) and suggested that consumers were limiting their expenditures to perceived necessities. Hence the marketing research division hypothesized that consumers might be viewing the school dental program as an additional, unnecessary expense.
- School enrollment in Iowa was declining, and the trend was expected to continue for several years, according to statistics and projections published by the Iowa Department of Public Instruction. Thus enrollment in the School Delta Dental Accident Program was likely to decline further.
- Employers were increasingly offering comprehensive health coverage (including dental coverage) to their employees through group medical insurance programs, according to U.S Department of Commerce publications and insurance industry publications. Furthermore, many Blue

analysis may even eliminate the need for conclusive research. Research in Use 4.1 illustrates how one company was able to make a final decision concerning a newly introduced service on the basis of secondary data obtained from internal company records and external data sources.

Case Study Method

Case Study *An in-depth examination, using a fairly flexible format, of a unit of interest such as a customer, store, salesperson, firm, or market area.*

A ***case study*** is an in-depth examination of a unit of interest. The unit can be a customer, store, salesperson, firm, market area, and so on. Julian Simon succinctly describes the purpose and nature of a case study as follows:

> It is the method of choice when you want to obtain a wealth of detail about your subject. You are likely to want such detail when you do not know exactly what you are looking for. The case study is therefore appropriate

Research in Use 4.1 *Continued*

Cross and Blue Shield group insurance contracts already included dental coverage for accidental injuries, according to internal company records. Therefore the demand for a separate dental-coverage program was bound to be weak.

- Mutual of Omaha, a leading firm in the insurance field, was offering a similar dental program at slightly lower cost, as indicated in their literature and brochures. Thus the dental program offered by Blue Cross and Blue Shield did not appear to have any unique competitive advantage.
- The School Delta Dental Accident Program was barely profitable, and even a slight increase in administrative costs would result in the program incurring a loss, according to cost and revenue data from internal company records.

All these insights, which emerged from a variety of secondary-data sources, were clearly unfavorable to the School Delta Dental Accident Program. Moreover, an informal investigation of key officials in 10 schools suggested that any change in the marketing strategy for the program was unlikely to arrest the decline in enrollment. Rather than conduct additional research, the marketing research division came to the following conclusion: "Based on the results . . . it appears that the School Dental Program is faced with a changing market. While alternative marketing strategies such as offering a reduced price for two or more children enrolling per family, . . . or consideration of new packaging strategies may be viable alternatives, it is unlikely that they will be successful." The company phased out the school dental program shortly after the marketing research division submitted its report.

Source: This example is based on "Product Analysis of the School Delta Dental Accident Program," a report prepared (in 1982) by the Marketing Research Division of Blue Cross and Blue Shield of Iowa, 636 Grand Avenue, Des Moines, Iowa 50307.

when you are trying to find clues and ideas for further research; in this respect, it serves a purpose similar to the clue-providing function of expert opinion [or the key-informant technique].

The specific method of the case study depends upon the mother wit, common sense, and imagination of the person doing the case study. The investigator makes up his procedure as he goes along, because he purposely refuses to work within any set categories or classifications; if he did so, he would not be obtaining the benefits of the case study.[3]

---

[3] Julian L. Simon, *Basic Research Methods in Social Science* (New York: Random House, 1969), p. 276. For a discussion of the potential advantages and drawbacks of the case study method, see also Thomas V. Bonoma, "Case Research in Marketing: Opportunities, Problems, and a Process," *Journal of Marketing Research*, 22 (May 1985), pp. 199–208.

The case study method, by virtue of its insight-generating potential, is a useful form of exploratory research. Let us look at a hypothetical example to illustrate the use of case studies in a marketing context.

> **Example.** Allied Associates Company (AAC) is a national chain of discount stores, with over 500 retail outlets spread across the country. During the past several years the company's sales have grown rapidly, and its profitability has remained above the industry average. Top management wants to identify the key elements crucial to the company's success and capitalize on these elements. Where and how should one start looking to accomplish top management's objectives?

The AAC scenario fits a research setting likely to benefit from the case study method; that is, it is a setting in which the company has a general research objective but is unsure of exactly what it is looking for. Moreover, each AAC retail outlet can be viewed as a case. Valuable insights can be gained by thoroughly examining certain selected retail outlets. For instance, the three best-performing and the three worst-performing stores can be studied in depth on numerous dimensions. Store size and layout, product lines carried, employee morale, and trade area characteristics are but a few of the factors for which data can be collected. In a case study only the investigator's time and imagination limit the number and types of factors to examine.

Insights about factors critical to AAC's success can be gained by looking for similarities and differences between the two types of cases, namely, the best and the worst stores. For instance, if employee morale is higher in all three best stores than in any of the worst stores, employee morale is likely to be critical to AAC's success. In contrast, a factor such as product lines carried is likely to be less important, although not necessarily unimportant, to AAC's success if both types of stores carry the same product lines. Insights like these can help researchers single out key determinants of performance variations across AAC stores. Before devoting attention and resources to the critical factors identified by the case analysis, top management may want to verify the association between the factors and store performance through a focused examination of a larger number of AAC stores; this study would, of course, constitute conclusive research.

As the AAC example illustrates, the analysis of case data is nonquantitative and primarily involves numerous comparisons and contrasts of the data. It requires an alert investigator capable of recognizing even subtle differences across cases as well as possible relationships among factors within a case.

Conducting exploratory research is not limited to using one of the four methods described—talking with knowledgeable individuals, focus group interviews, analysis of secondary data, and case studies—although these are the most frequently used methods. Variations or combinations of these methods can also be employed in an exploratory-research project.

Insights gained through exploratory research pave the way for conclusive research. In fact, many research projects involve an exploratory phase followed by a conclusive phase. Two different kinds of conclusive research are discussed in the next section.

DESCRIPTIVE VERSUS EXPERIMENTAL RESEARCH

There are two basic forms of conclusive research: descriptive research and experimental research. The latter is also known as causal research. The distinction between the two is based on the primary purpose of a conclusive-research project and the nature of the inferences that can be drawn from it. The goal of ***descriptive research***, as the name implies, is essentially to describe something. Specifically, it is intended to generate data describing the composition and characteristics of relevant groups of units such as customers, salespeople, organizations and market areas. Data collected through descriptive research can provide valuable information about the study units along relevant characteristics and also about associations among those characteristics (more on this later).

Descriptive Research *A form of conclusive research that is intended to generate data describing the composition and characteristics of relevant groups of units such as customers, salespeople, organizations, or market areas.*

A drawback of descriptive research is that it generally cannot provide the type of evidence necessary to make *causal inferences* about relationships between variables. Some form of ***experimental research*** is necessary to make causal inferences with confidence. To be able to say that *X* (e.g., shelf space assigned to a product in a supermarket) has a causal influence on *Y* (e.g., unit sales of the product), one must gather data under controlled conditions—that is, holding constant, or neutralizing the effect of, all variables other than *X* capable of influencing *Y*, and systematically manipulating the levels of *X* to study its impact on *Y*. Manipulation of the presumed causal variable and control of other relevant variables are distinct features of experimental research. Data collected through experimental research therefore can provide much stronger evidence of cause and effect than can data collected through descriptive research.

Experimental Research *A form of conclusive research that is intended to generate the type of evidence necessary for confidently making causal inferences about relationships between variables.*

The preceding discussion does not necessarily mean that analysis of descriptive research data cannot *suggest* possible causal linkages between variables, especially when the effects of uncontrolled variables are filtered through certain analysis techniques available for that purpose. In fact, rather than view descriptive-versus-experimental research as a clear-cut dichotomy, we should think of conclusive projects as falling along a ***research continuum***, with "purely descriptive with no control" at one extreme and with "purely experimental with strict control and manipulation" at the other extreme. Virtually all real-life conclusive projects fall somewhere along this continuum, although the point where "descriptive" ends and "experimental" begins is subjective and somewhat arbitrary.

Research Continuum *A line or dimension along which conclusive research projects can be arrayed, with "purely descriptive with no control" at one extreme and "purely experimental with strict control and manipulation" at the other.*

For the sake of highlighting and illustrating key differences between

Table 4.3 *Association Between Customers' Income Levels and Perceptions*

Customers' Income Level	Number (%) of Customers Perceiving SNB as Conveniently Located	Not Conveniently Located	Total
Less than $30,000	450 (90%)	50 (10%)	500 (100%)
$30,000 or more	150 (30%)	350 (70%)	500 (100%)

the two types of conclusive research, we will focus for the time being on projects falling toward either end of the research continuum. In other words, in the following discussion we will use the terms *descriptive research* and *experimental research* cautiously, bearing in mind the preceding caveats concerning the descriptive-versus-experimental dichotomy.

Recall situation B presented earlier, pertaining to Saver's National Bank (SNB). The types of data SNB desires—the demographic profile of its customers as well as their perceptions about SNB—are descriptive data. Hence descriptive research is the appropriate form of conclusive research to be used in situation B.

Descriptive research of a relevant group of units can reveal the proportion of units falling into different categories of a variable. For instance, descriptive research using a survey of, say, 1000 SNB customers might provide the following types of information:

- Fifty percent of SNB's customers have annual family incomes of less than $30,000, while the remaining 50% have annual family incomes of $30,000 or more.
- The bank's location is perceived to be convenient (and hence a strength) by 60% of its customers, while it is perceived to be inconvenient (and hence a weakness) by the remaining 40%.

Descriptive research can also point out associations between variables. For instance, in the SNB scenario simultaneous analysis of data pertaining to customer income levels and perceptions about location may reveal that while 90% of the customers with income levels less than $30,000 perceive SNB's location to be convenient, only 30% of the customers with income levels of $30,000 or more perceive SNB as conveniently located. In other words, customer income levels are apparently associated with their perceptions about SNB's location, as illustrated in Table 4.3. Customers with lower income levels are more likely to perceive SNB as being conveniently located. By contrast, higher-income customers are less likely to perceive SNB as being conveniently located.

Descriptive research can provide profile descriptions (e.g., the proportion of high- and low-income customers in SNB's clientele) and point out associations between profile characteristics (e.g., the association between customer income levels and perceptions about locational convenience). But it cannot establish cause and effect between characteristics of interest. The strong association between customers' income levels and perceptions of SNB's locational convenience does not imply that the former is the *cause* of the latter. On the basis of descriptive data, to conclude a causal linkage between higher-income levels and unfavorable perceptions about location will be erroneous—and in SNB's case even a bit ridiculous. For instance, most of SNB's higher-income customers may reside in areas far away from SNB. If so, the cause of perceptions about SNB's location may be the relative locations of customer residences rather than customer income levels. The analysis of data on income levels and perceptions did not control for the possible influence of the location of customer residences vis-à-vis SNB's location. The dangers of making causal inferences based on cross-tabulations of descriptive data (such as in Table 4.3) are discussed in detail in Chapter 16. To make causal inferences with confidence, one must conduct experimental research, as described later in the chapter. General approaches used in conducting descriptive research are discussed next.

CONDUCTING DESCRIPTIVE RESEARCH

Descriptive research, or research falling closer to the descriptive end than to the experimental end of the research continuum, is by far the most frequently used form of conclusive research. On the basis of the general data collection approach used, descriptive-research studies are classified into two basic types: cross-sectional studies and longitudinal studies.

Cross-Sectional Study *A one-time study involving data collection at a single period in time.*

Longitudinal Study *A repeated measurement study involving data collection at several periods in time.*

Cross-sectional studies are onetime studies involving data collection at a single period in time. ***Longitudinal studies*** are repeated-measurement studies involving data collection at several periods in time. While a cross-sectional study yields a "snapshot" of a situation being researched, a longitudinal study produces a "motion picture" (or a series of snapshots) of a situation over time. In general, longitudinal studies are more informative than cross-sectional studies, just as motion pictures are more revealing than still pictures. But longitudinal studies are also more expensive. Further, the choice between the two depends on the objectives of the research.

Cross-Sectional Sample *A group of units selected specifically and solely for data collection at a single period in time.*

Panel *A fixed group of units recruited to provide measurements over a period of time.*

A cross-sectional study makes use of a ***cross-sectional sample*** or group of units (e.g., consumers, stores, organizations) selected specifically and solely for the onetime data collection. The sample is disbanded after the data are collected. A longitudinal study typically employs a ***panel*** or group of units recruited to provide measurements over a period of time.

At the conclusion of each measurement phase, a panel is maintained intact for future use.

Thus there is a clear distinction between a cross-sectional sample and a panel. However, the distinction is blurred in practice. Sometimes, cross-sectional samples are used in longitudinal studies, and panels (or certain segments of panels) are used in cross-sectional studies.

For example, suppose the U.S. Department of Defense wants to ascertain people's attitudes toward nuclear weapons and nuclear warfare and to monitor changes in those attitudes over time. Specifically, suppose the Defense Department wants to measure the public's attitudes on a quarterly basis. The appropriate research design is clearly a longitudinal study. However, one need not necessarily set up a panel of respondents for this purpose. Rather, one can select a separate cross-sectional sample of people for each quarterly measurement. In fact, using cross-sectional samples instead of a panel of respondents may have advantages, as will be discussed later. The point is that a cross-sectional study and the use of a cross-sectional sample (or a longitudinal study and the use of a panel) are not paired in all situations.

Cross-Sectional Studies

We saw earlier that descriptive research is the most frequently used form of conclusive research. The most popular method within the domain of descriptive research is the cross-sectional study. Also, cross-sectional studies account for the majority of a formal research projects involving primary-data collection.

By definition, a cross-sectional study involves data collection at only one period in time. However, data pertaining to different periods in time can be obtained through such a study. In other words, the *scope* of the data collected is not necessarily limited to the time at which a cross-sectional study is conducted. The following example clarifies this point.

Example. Capital Cola Company (CCC) wants to obtain a descriptive profile of consumers who drink only caffeine-free soft drinks. Specifically, CCC wants to know the demographic composition and use characteristics (e.g., brands consumed, volume consumed, frequency of consumption) of such consumers. The company is considering a cross-sectional study involving a random sample of consumers.

Can CCC obtain data other than current demographic and use characteristics through the cross-sectional study? Yes, it can. For instance, the study respondents can be asked about their *past* consumption of soft drinks. What brand(s) did they consume during the past year? Was their past consumption more than, about the same as, or less than their consumption at the present time? Similarly, data can also be obtained about *future* consumption: Do the respondents intend to continue buying their present brand(s) of soft drinks? Do they plan to change their volume of consumption during the coming year? Answers to questions like these can be valuable to CCC in understanding the dynamics of soft-drink consumption patterns.

The CCC example implies that a cross-sectional study is capable of yielding longitudinal data of sorts, that is, data pertaining to an interval of time rather than a single point in time. Indeed, many cross-sectional studies collect longitudinal data of this nature. However, a rather serious limitation of such data is that their accuracy depends heavily on the mental capabilities (memory of past events and intentions about future behavior) of the study respondents. For the most part, consumers' memories are unreliable, particularly with respect to remembering things that occurred very far in the past. Hans Zeisel makes this point clearly:

> The memory of a minor expenditure, or a fleeting observation will disappear within days. Moreover, to rely on memory is always a treacherous undertaking, because of unconscious and even conscious forces that tend to distort it. The question, for instance, "For whom did you vote?" asked one day after election day, will always elicit too many votes for the victorious candidate.[4]

Likewise, consumer intentions may be a poor indicator of future behavior.[5] This problem becomes increasingly severe as the time frame extends further into the future. A longitudinal study is much more reliable than a cross-sectional study for monitoring changes over time. A longitudinal study is more reliable because it relies less on consumers' mental capabilities through more frequent monitoring of events as close to their time of occurrence as feasible.

The sample used in a cross-sectional study is typically selected solely for the onetime data collection and disbanded after the data collection. However, several firms, especially certain commercial marketing research firms, maintain ***omnibus panels*** as a source of samples for cross-sectional studies. Firms maintaining omnibus panels have the capability to custom-design cross-sectional samples to meet specific research requirements. Such samples are composed of panel members who are returned to the panel after participating in a cross-sectional study.

Omnibus Panel *A panel that serves as a ready source of samples for a variety of cross-sectional studies over time.*

One firm that maintains an omnibus panel is National Family Opinion, Inc. (NFO). The NFO panel consists of 130,000 households in the United States and represents the population on a variety of characteristics. According to NFO:

---

[4] Hans Zeisel, *Say It with Figures* (New York: Harper & Row, 1968), p. 202; see also Yoram Wind and David Lerner, "On the Measurement of Purchase Data: Surveys Versus Purchase Diaries," *Journal of Marketing Research,* 16 (February 1979), pp. 39–47, for a study that emphasizes the pitfalls of surveys relying on respondents' memories.

[5] Studies have shown that many consumer purchase decisions are made on impulse. For instance, Kenneth E. Runyon, *Consumer Behavior and the Practice of Marketing* (Columbus, Ohio: Merrill, 1977), p. 324, has reported that about half of all buying decisions in supermarkets are impulse decisions made on the spot. The point is that inferring behavior from intentions is very dangerous.

> Fifty demographics on each household, ranging from income, education and occupation to ownership of such things as video cassettes, are on file to provide easy computer selection. . . . From this background we can economically produce cross-sectional samples of most segments of the consumer market (e.g., farmers, large-city dwellers, homeowners, high income households, racial minorities, singles, pet owners, etc.).[6]

The NFO print advertisement shown in Exhibit 4.1 illustrates how a custom-designed cross-sectional sample can be chosen from an omnibus panel.

The Home Testing Institute (HTI) is another firm specializing in omnibus panels. The HTI maintains several unique panels, such as the HTI baby panel. Exhibit 4.2 describes this panel and lists the types of cross-sectional samples that can be chosen.

Longitudinal Studies

The primary purpose of longitudinal studies is to monitor changes over time. The following examples illustrate situations calling for longitudinal studies:

- A firm manufacturing frozen orange juice is interested in keeping track of consumers' use of brands and types of orange juice.
- A social service organization wishes to monitor the well-being of an urban community.
- A public health institution wants to ascertain whether and how people's eating and drinking habits vary from season to season.

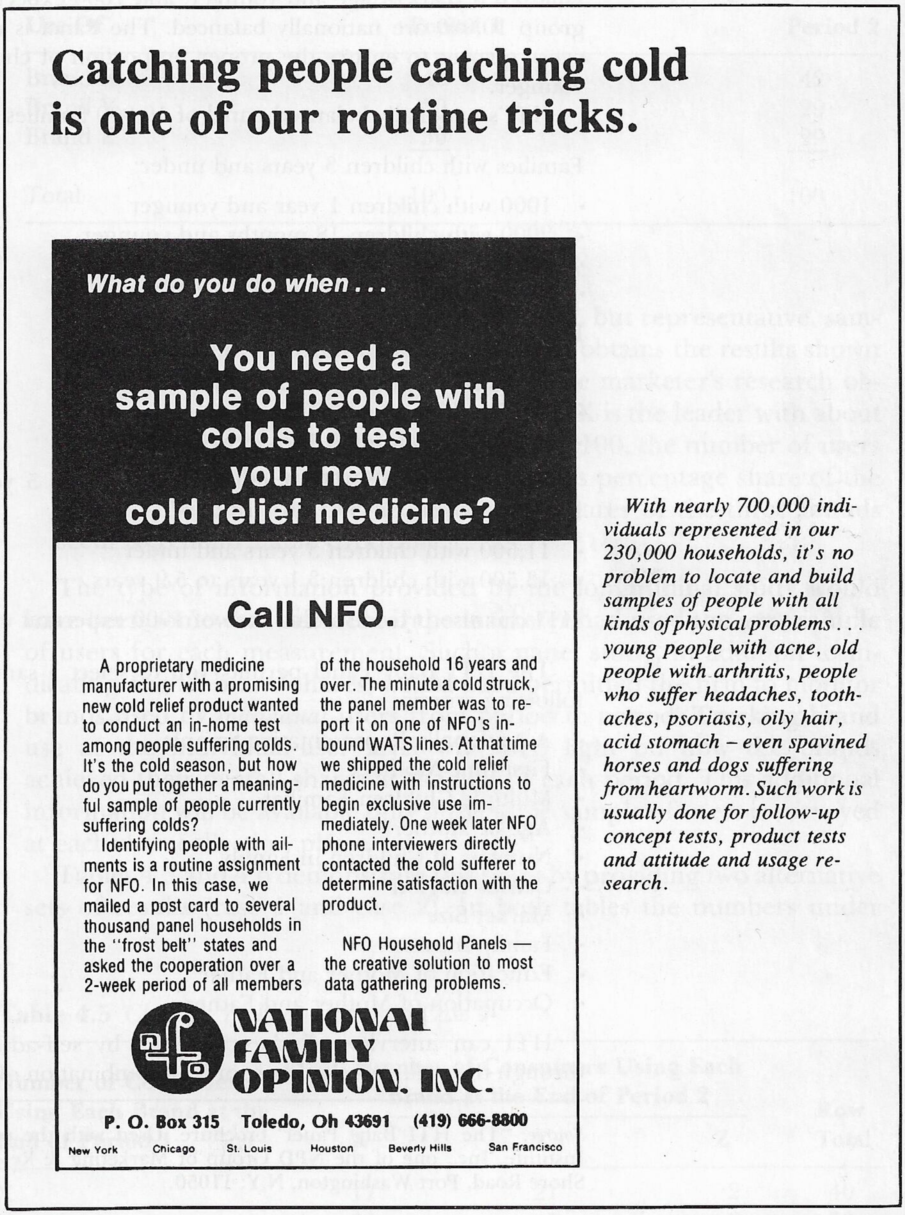
Successive measurements in longitudinal studies can be obtained from a physically different, but representative, sample of units or from the same sample of units each time. Although both sample options will yield longitudinal data, the nature of the findings and the implications can differ, as demonstrated in the next example.

Example. The marketer of brand X of a product wishes to monitor X's share versus competing brands. For the sake of simplicity, let us assume there are only two other brands competing with brand X, say, brand Y and brand Z. At the end of each suitable period (e.g., month or quarter), the marketer can survey a different cross section of users of the product and ascertain the brands they use. Alternatively, the marketer can form a panel of users and ask the panel members to periodically report the brands they use; in this case the *same* respondents will provide data during the course of the study.

---

[6] "Creative Use of the NFO Household Panel," a brochure prepared by National Family Opinion, Inc., 660 Madison Avenue, New York, N.Y. 10021.

Exhibit 4.1 *Selecting a Special Cross-Sectional Sample from an Omnibus Panel*



Catching people catching cold is one of our routine tricks.

What do you do when . . .

You need a sample of people with colds to test your new cold relief medicine?

Call NFO.

A proprietary medicine manufacturer with a promising new cold relief product wanted to conduct an in-home test among people suffering colds. It's the cold season, but how do you put together a meaningful sample of people currently suffering colds?

Identifying people with ailments is a routine assignment for NFO. In this case, we mailed a post card to several thousand panel households in the "frost belt" states and asked the cooperation over a 2-week period of all members of the household 16 years and over. The minute a cold struck, the panel member was to report it on one of NFO's inbound WATS lines. At that time we shipped the cold relief medicine with instructions to begin exclusive use immediately. One week later NFO phone interviewers directly contacted the cold sufferer to determine satisfaction with the product.

NFO Household Panels — the creative solution to most data gathering problems.

With nearly 700,000 individuals represented in our 230,000 households, it's no problem to locate and build samples of people with all kinds of physical problems . . . young people with acne, old people with arthritis, people who suffer headaches, toothaches, psoriasis, oily hair, acid stomach – even spavined horses and dogs suffering from heartworm. Such work is usually done for follow-up concept tests, product tests and attitude and usage research.

NATIONAL FAMILY OPINION, INC.

P. O. Box 315 Toledo, Oh 43691 (419) 666-8800

New York Chicago St. Louis Houston Beverly Hills San Francisco

Source: "Creative Use of the NFO Household Panel," a brochure prepared by National Family Opinion, Inc., 660 Madison Avenue, New York, N.Y. 10021. Reprinted by permission.

Exhibit 4.2 *Description of the HTI Baby Panel*

HTI's innovative Baby Panel consists of 35,000 cooperating families with children 5 years of age and younger, and 1000 expectant mothers. Among this group 10,000 are nationally balanced. The Panel is updated demographically every quarter to ensure the proper proportion of children 5 years of age and younger.

HTI's nationally balanced panel of 10,000 families includes:

Families with children 3 years and under:

- 1000 with children 1 year and younger
- 2000 with children 18 months and younger
- 3500 with children 24 months and younger
- 6000 with children 36 months and younger

Families with children 3 to 5 years of age:

- 1500 with children 3.1 to 4 years
- 1250 with children 4.1 to 5 years
- 1250 with children 5.1 to 5.9 years

The remaining 25,000 families with children 5 years of age and younger include:

- 11,500 with children 3 years and under
- 13,500 with children 3.1 years to 5.9 years

HTI can also provide a data base of 1000 expectant mothers.

The HTI Baby Panel permits you to select a sample based on any of the following criteria:

- A Nationally Representative Sample
- First-time Mothers
- Multiple Children Families
- Age of Mother
- Number of Children in Family
- Census Region
- Market Size
- Family Income
- Education of Mother and Father
- Occupation of Mother and Father

HTI can interview all these samples by self-administered questionnaires through the mail, by telephone, or by a combination of these survey techniques.

Source: "The HTI Baby Panel" brochure. Used with the permission of Home Testing Institute, Inc., one of the NPD Group of Marketing & Research Companies, 900 West Shore Road, Port Washington, N.Y. 11050.

Table 4.4 *Results of Longitudinal Brand Use Study*

Use Of	Number of Consumers Using Brand at the End of Period 1	Period 2
Brand X	40	42
Brand Y	30	29
Brand Z	30	29
Total	100	100

Suppose the marketer surveys a different, but representative, sample of 100 users for each measurement and obtains the results shown in Table 4.4. Two insights pertaining to the marketer's research objective emerge from Table 4.4. First, brand X is the leader with about a 40% brand share (since the sample size is 100, the number of users of a brand can be interpreted as the brand's percentage share of the market). Second, the overall percentage shares of the three brands seem to remain stable over time.

The type of information provided by the longitudinal study would have been somewhat different if the marketer had used the *same* sample of users for each measurement. Such a panel study, in addition to indicating overall brand shares, would have permitted the firm to monitor brands used by *individual* users from period to period. Tracking brand use at the individual-user level will shed light on how the brands achieved their overall shares at the end of each period. This additional information will be available only if the same sample of users is surveyed at each data collection phase.

Tables 4.5 and 4.6 demonstrate this point by providing two alternative sets of results (case 1 and case 2). In both tables the numbers under

Table 4.5 *Changes in Brand Shares: Case 1*

Number of Consumers Using Each Brand at the End of Period 1	Number of Consumers Using Each Brand at the End of Period 2 X	Y	Z	Row Total
X	17	21	2	40
Y	23	5	2	30
Z	2	3	25	30
Column Total	42	29	29	100

Table 4.6 *Changes in Brand Shares: Case 2*

Number of Consumers Using Each Brand at the End of Period 1	Number of Consumers Using Each Brand at the End of Period 2			
	X	Y	Z	Row Total
X	38	1	1	40
Y	2	27	1	30
Z	2	1	27	30
Column Total	42	29	29	100

"Row Total" represent brand shares at the end of period 1 and the numbers under "Column Total" represent brand shares at the end of period 2. The overall brand shares at the end of periods 1 and 2 in Tables 4.5 and 4.6 are identical to those in Table 4.4. Hence the inferences we made earlier about the overall shares of brands X, Y, and Z still hold. However, Tables 4.5 and 4.6 contain additional data about how each brand obtained its overall share. To illustrate, consider the first row of numbers in Table 4.5. These numbers reveal that, of the 40 consumers who bought brand X in period 1, 17 bought brand X again in period 2, 21 switched to brand Y, and 2 switched to brand Z. The first row of numbers in Table 4.6 reveals that, of the 40 consumers who bought brand X in period 1, 38 bought brand X again in period 2, 1 switched to brand Y, and 1 switched to brand Z. The other two rows in each table can be interpreted in similar fashion.

According to Table 4.5 (case 1), while all three brands have stable shares, only brand Z appears to enjoy *brand loyalty;* a majority of brand X and brand Y users switched brands during each period. In contrast, Table 4.6 (case 2) shows all three brands' having stable shares and enjoying high brand loyalty. Clearly, the marketing implications for brand X emerging from case 1 are quite different than those emerging from case 2.

True Panel *A panel participating in a longitudinal study involving measurement of the same sample units, with respect to the same variables, over time.*

The point of this discussion is that a longitudinal study using the same sample of respondents will provide richer information than one using a series of different samples. Indeed, the dynamics of changes between measurements can only be captured by using the same panel of respondents. Such a panel has been labeled a ***true panel***, to distinguish it from omnibus panels used to generate different cross-sectional samples at various periods in time.[7]

---

[7] Gilbert A. Churchill, Jr., *Marketing Research: Methodological Foundations* (Chicago: Dryden Press, 1987), pp. 85–91, provides an excellent discussion of the additional capabilities of a true panel over those of an omnibus panel.

A true-panel study, compared with a longitudinal study using different samples for the various measurements, is also capable of generating more data directly pertaining to the research purpose, for the following reasons:

1. A true panel is a captive sample of willing respondents who are likely to tolerate extended interviews or fill out lengthy questionnaires.
2. Background data such as demographic and life-style data need not be collected from panel respondents during each measurement. Therefore, for a given interview or questionnaire length, more data of primary research interest can be collected.[8]

Consumer panels are not without drawbacks. A major difficulty in setting up a panel is identifying a *representative* sample of respondents who are willing to cooperate over a long period of time. Certain types of consumer groups are especially hard to recruit for panels. Examples of such groups include nonwhites, housewives under 25, employed housewives, and illiterates.[9] Even if a researcher initially succeeds in putting together a representative panel, there will be some attrition of panel members during a longitudinal study. As a result, the panel may no longer be representative unless the departing members are replaced with similar new members—an expensive and time-consuming task.

Another potential problem with consumer panels is that the multiple-survey participation by panel members may, over a period of time, induce them to alter their natural or usual behavior. Sudman and Ferber label this difficulty ***panel conditioning*** and illustrate it as follows:

Panel Conditioning *A bias that occurs when multiple-survey participation by panel members, over a period of time, induces them to alter their natural or usual behavior.*

> A homemaker who keeps a weekly purchase diary with a separate section for canned goods may feel guilty about not having made any such purchases recently and may deliberately purchase some canned goods to be able to record it in the diary. In a similar fashion, a family asked month after month about ownership of savings accounts may decide to open a savings account, even though they originally had no such intention.[10]

Most commercial marketing research firms operating consumer panels take steps to ensure that the behavior of panel members is normal and that the reporting of such behavior is accurate. For instance, the National Purchase Diary Panel, Inc. (NPD), which maintains a panel of families, states:

---

[8] For a complete discussion of the advantages and the limitations of panels, see Seymour Sudman and Robert Ferber, *Consumer Panels* (Chicago: American Marketing Association, 1979); see also Simon, *Basic Research Methods,* Chap. 19, and Zeisel, *Say It with Figures,* Chap. 13.

[9] "Sharpening Marketing Decisions with Diary Panels," a brochure prepared by National Purchase Diary Panel, Inc., 900 West Shore Road, Port Washington, N.Y. 11050, p. 20.

[10] Sudman and Ferber, *Consumer Panels,* p. 6.

> Families are dropped from the panel at their request or if they fail to return three of their last five diaries. In addition NPD has instituted the policy of eliminating family members after 4–5 years of reporting since buying behavior may change after detailed reporting of purchases over this period of time.[11]

Obviously, precautionary procedures like the ones used by NPD add to the expense of maintaining and operating a panel. Nevertheless, they are necessary if panel data are to be trusted.

Because of the potential limitations of true panels, researchers may be wise to restrict their use to situations where periodic monitoring of the same respondents is essential. The unique capabilities of a true panel will not be needed when the purpose of a longitudinal study is merely to track variables at a macro, or aggregate, level. Under those circumstances choosing several cross-sectional samples from the general population—or from a suitable omnibus panel if a special type of sample is required—may be better than using a true panel.

CONDUCTING EXPERIMENTAL RESEARCH

Experimental research is intended to generate the type of evidence necessary for confidently making causal inferences about relationships between variables. It is designed to overcome the lack of control suffered by descriptive-research projects. It involves data collection in an environment wherein the hypothesized causal variable is manipulated and the effects of other relevant variables are controlled. To illustrate, consider a consumer goods firm that wants to determine the impact of advertising on sales. To accomplish this objective, it can proceed as follows:

1. Select a group of distinct market areas that have similar demographic, socioeconomic, and competitor characteristics.
2. Vary the level of advertising expenditure from market to market, keeping all other marketing variables, like price and promotion, constant.
3. Monitor sales over a sufficient length of time.
4. Analyze the data to see whether the pattern of variation in sales across markets is consistent with the pattern of variation in advertising expenditures.

These steps constitute an experimental- rather than a descriptive-research approach because factors other than advertising that can influence sales are, by and large, held constant. Assuming that no major differences in external conditions occur within the group of markets during the research period—differences such as a sudden change in the unemployment rate or competitive activity in some markets but not in others—variations in sales can be attributed to those in advertising expenditures.

---

[11] "Sharpening Marketing Decisions," p. 20.

In contrast, data on advertising and sales gathered through a descriptive-research project—such as a survey of a representative sample of consumer goods firms asking them to indicate their advertising and sales expenditures during some past period—are not capable of providing as strong an indication of cause and effect. A positive correlation between advertising and sales revealed by an analysis of the survey data cannot be interpreted to mean that higher advertising will lead to higher sales for two reasons. First, although high advertising expenditures and high sales apparently occurred together, the latter might be due to other factors such as the larger sizes or more favorable competitive positions of firms reporting higher sales. Second, the sample firms might have allocated a certain portion of their past sales or anticipated sales to advertising, as is frequently done in practice; if so, sales levels determine advertising levels rather than vice versa.

Planned manipulation of the causal variable (e.g., advertising) and effective control of other variables are preconditions for making causal inferences with confidence. Another precondition is that the causal variable (advertising) and effect variable (sales) must occur in the proper time sequence. To ascertain whether advertising causes sales, for instance, one must change advertising levels first, without their being influenced by anticipated sales; any changes in sales must be measured later. These preconditions distinguish experimental research from descriptive research. They are discussed at length in Chapter 8.

To further illustrate how experimental research is conducted, consider the TEA example described in situation C at the beginning of the chapter. The president of TEA wants to determine the impact on sales revenues and profits when price is manipulated (price is reduced by 15%). An experimental-research study can be set up for situation C, as outlined next.

The eight communities having TEA restaurants have similar characteristics. Hence we can arbitrarily divide the eight restaurants into two equivalent groups, say, group A and group B. Prices are dropped by 15% in the group A restaurants but maintained at current levels in the group B restaurants. After a few weeks sales revenues and profits are compared for groups A and B. Any significant differences between the two groups will be an indication of a causal impact of the price reduction on sales revenues and profits. Assuming all other factors capable of influencing sales or profits remained at similar levels in groups A and B during the research, any significant difference in sales revenues and/or profits between the two groups can be unambiguously attributed to the price reduction.

The preceding design is a two-group experimental design involving an *experimental group* (A) and a *control group* (B). Since the two groups are equivalent except for the change in the causal variable (e.g., price reduction) in group A, the control group captures all influences on the effect variable (e.g., sales or profits) other than that of the causal variable.

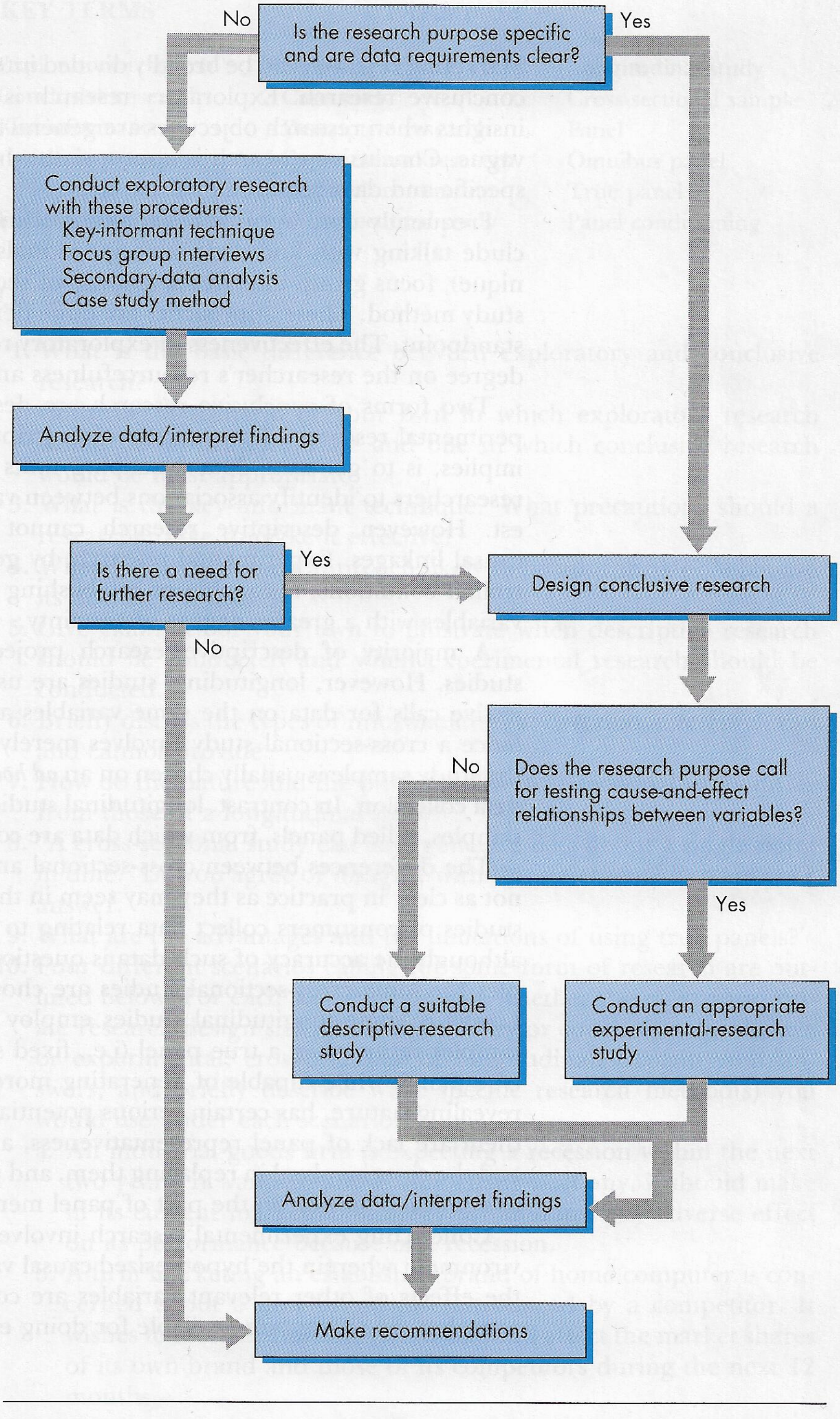
The following alternative approach can be used if dividing the eight restaurants into equivalent experimental and control groups is not feasible. Sales revenues and profits of the eight restaurants are monitored for several weeks. Prices are then dropped by 15% in all eight restaurants, but other marketing variables are left unchanged. Sales revenues and profits of the eight restaurants are again monitored for the same number of weeks as before. Any significant differences between the "before" and "after" measurements will shed light on the causal impact of the price reduction on sales revenue and profits. This approach is known as a one-group, before-after experimental design. This and the preceding design, as well as several other experimental designs, are covered in detail in Chapter 9.

DETERMINING WHICH TYPE OF RESEARCH TO CONDUCT

Choosing the most appropriate type of research—exploratory or conclusive—in a situation is somewhat subjective. The choice depends not only on the nature of the situation but also on how the decision maker and researcher perceive it. Exploratory research is most appropriate in situations where the research objectives are general and data requirements are unclear. Some form of conclusive research is appropriate in other situations. Insights gained through exploratory research typically form the foundation for more formal conclusive research. Occasionally, however, exploratory-research results may strongly suggest that further research of a conclusive nature may be unnecessary or unproductive.

In situations calling for conclusive research, the choice of the type of conclusive research—descriptive or experimental—depends on whether testing causal relationships between variables is the primary research purpose. If so, some form of experimental research is appropriate; if not, descriptive research will suffice. Of course, as implied in preceding sections of this chapter, the descriptive-versus-experimental distinction exists more along a research continuum than as a clear-cut dichotomy. Therefore one may at times be able to make tentative causal statements on the basis of data from a well-designed and -conducted descriptive-research study. By the same token, one can rarely conduct a pure experimental-research study. As we will see in Chapters 8 and 9, many experimental-research studies suffer from varying degrees of lack of control owing to resource, environmental, and other constraints. Exhibit 4.3 is a flow diagram that offers general guidelines for identifying the most appropriate research type(s) to employ in a situation calling for research.

Exhibit 4.3 *Flow Diagram for Selecting the Appropriate Research Type*


No
Is the research purpose specific and are data requirements clear?
Yes
Conduct exploratory research with these procedures:
Key-informant technique
Focus group interviews
Secondary-data analysis
Case study method
Analyze data/interpret findings
Is there a need for further research?
Yes
No
Design conclusive research
No
Does the research purpose call for testing cause-and-effect relationships between variables?
Yes
Conduct a suitable descriptive-research study
Conduct an appropriate experimental-research study
Analyze data/interpret findings
Make recommendations

SUMMARY

Marketing research can be broadly divided into exploratory research and conclusive research. Exploratory research is useful for gaining initial insights when research objectives are general and data requirements are vague. Conclusive research is appropriate when research objectives are specific and data requirements are clear.

Frequently used approaches for conducting exploratory research include talking with knowledgeable individuals (the key-informant technique), focus group interviews, analysis of secondary data, and the case study method. These approaches are quite flexible from the researcher's standpoint. The effectiveness of exploratory research depends to a large degree on the researcher's resourcefulness and skills.

Two forms of conclusive research are descriptive research and experimental research. The purpose of descriptive research, as the name implies, is to generate data describing units of interest and to enable researchers to identify associations between variables or factors of interest. However, descriptive research cannot unambiguously establish causal linkages. Experimental research, by generating data under controlled conditions, is capable of establishing cause and effect between variables with a greater degree of certainty.

A majority of descriptive-research projects involve cross-sectional studies. However, longitudinal studies are used when the research objective calls for data on the same variables at different points in time. Since a cross-sectional study involves merely a onetime measurement, the study sample is usually chosen on an *ad hoc* basis and disbanded after data collection. In contrast, longitudinal studies typically use permanent samples, called panels, from which data are collected on a periodic basis.

The differences between cross-sectional and longitudinal studies are not as clear in practice as they may seem in theory. Many cross-sectional studies of consumers collect data relating to past and future behavior, although the accuracy of such data is questionable. Moreover, the samples for some cross-sectional studies are chosen from omnibus panels. Likewise, some longitudinal studies employ a series of cross-sectional samples rather than a true panel (i.e., fixed sample) of respondents. A true panel, while capable of generating more data and data of a more revealing nature, has certain serious potential limitations. Chief among them are lack of panel representativeness, attrition of panel members and the costs involved in replacing them, and panel conditioning leading to atypical behavior on the part of panel members.

Conducting experimental research involves collecting data in an environment wherein the hypothesized causal variable is manipulated and the effects of other relevant variables are controlled. A variety of approaches, or *designs,* are available for doing experimental research.

KEY TERMS

Exploratory research
Conclusive research
Key-informant technique (expert-opinion survey)
Focus group interview
Secondary data
Case study
Descriptive research
Experimental research
Research continuum
Cross-sectional study
Longitudinal study
Cross-sectional sample
Panel
Omnibus panel
True panel
Panel conditioning

QUESTIONS

1. What is the basic difference between exploratory and conclusive research?
2. Construct a scenario of your own in which exploratory research would be most appropriate and one in which conclusive research would be most appropriate.
3. What is the key-informant technique? What precautions should a researcher take to make it effective?
4. When is the case study approach likely to be appropriate? What are its distinct features?
5. Give examples of your own to illustrate when descriptive research should be conducted and when experimental research should be conducted.
6. Briefly discuss the types of information that descriptive research can and cannot provide.
7. How do the nature and the purpose of a cross-sectional study differ from those of a longitudinal study?
8. "A cross-sectional study can only generate data about a single point in time." Do you agree or disagree with this statement? Explain your answer.
9. What are the advantages and the limitations of using true panels?
10. Four different scenarios calling for some form of research are outlined below. For each scenario, indicate whether the most appropriate research design should be exploratory or conclusive; descriptive or experimental; cross-sectional or longitudinal. Explain your answers, and briefly describe what specific research method(s) you would use under each scenario.
    a. An industrial-goods firm is expecting a recession within the next two years. It wants to know what changes, if any, it should make in its current marketing strategy to minimize any adverse effect on its performance because of a recession.
    b. A firm marketing an established brand of home computer is concerned about a new brand just introduced by a competitor. It wishes to monitor how the new brand will affect the market shares of its own brand and those of its competitors during the next 12 months.

c. A packaged-foods firm has developed a new frozen dinner it feels is superior to other frozen dinners already on the market. It wants to develop a unique promotional theme for the new dinner to set it apart from competing brands and to appeal to a broad cross section of consumers.

d. A large bank currently has over twenty thousand loan customers. It is concerned about the increasing number of customers who are defaulting on their loans. It therefore wants to know whether customers with good payment records differ from those who defaulted recently. Specifically, the bank wants information on each of the following characteristics: age, sex, income, occupation, marital status, and past credit record.

Cases for Part Two

CASE 2.1 *BETTCHER INDUSTRIES, INC.*

CASE 2.2 *FAST-TEST: TO USE OR NOT TO USE?*

CASE 2.3 *SUPERIOR LANDSCAPE COMPANY*

CASE 2.1

Bettcher Industries, Inc.

Dan Buttner, marketing research manager for Bettcher Industries, Inc., was wondering how he should respond to a letter he had just received from a marketing professor. In that letter the professor had asked Mr. Buttner if he could provide, from his experience, examples of real-life applications of marketing research for possible use in marketing research classes. Here are some excerpts from the professor's letter:

> One of the greatest challenges that I continue to face as an instructor of marketing research is to make students appreciate the reality of marketing research. Surveys of professionals like you have consistently shown that fresh college graduates are too idealistic and technique-oriented; that is, they are not fully aware of the *practical* limitations of the techniques and tools that they learn in the classroom.
>
> I am sure you would agree with me that a fruitful way of attempting to remedy this weakness in students is to expose them to *real-life* applications of marketing research. Unfortunately, unlike examples of different types of advertising or packaging, examples of actual applications of marketing research are hard to come by.
>
> Can you please send me any material(s), of a nonconfidential nature, that can help students gain an understanding of marketing research practice?

Bettcher Industries, Inc., located in Birmingham, Ohio, was founded in 1944 and incorporated in 1955. Bettcher designs and manufactures

---

Information for this case was provided by courtesy of Bettcher Industries, Inc., State Route 60 and Ohio Turnpike, Birmingham, OH 44089.

processing and portion control equipment for meat, fish, and poultry. The company's major product line is a wide range of machines for pressing, forming, tenderizing, slicing, and portioning meats and related products. Markets for these machines include meat-processing plants, restaurants, hotels, supermarkets, caterers, and all major food service organizations and food processors.

Bettcher holds a position of leadership in the meat industry and has a reputation for introducing highly innovative products. Such products include a line of hand-held Whizard brand trimmers. These air- and electric-powered devices can replace straight knives used in meat plants throughout the United States and foreign nations. The company is also noted for its Whizard brand protective wear, a line of accessories, such as safety gloves and arm protectors, of cut-resistant fabric for use by meat processors, food stores, and general industrial/commercial food processors. According to a company brochure, these gloves are made of a "space-age, seamless, knitted fabric," which consists of a stainless steel core encased in soft but high-strength textile. Bettcher has sold Whizard Protective Wear to several hundred organizational customers in a variety of industries. These customers include institutions such as John Deere, Levi Strauss, GE, GM, Coors, IBM, Exxon, Safeway, RCA, Rath, Land O Lakes, the U.S. Postal Service, the Texas Department of Corrections, Firestone, and J. C. Penney.

When Mr. Buttner received the professor's request, Bettcher Industries had just signed an agreement to acquire Sam Stein Associates of Sandusky, Ohio. Stein designs and manufactures batter and breading equipment for preportioned meat, fish, and poultry. It also makes fryers, cooking ovens, and snack-food-processing equipment. The product lines of the two companies were believed to complement each other and appeal to similar markets. While Bettcher equipment is used from the point of slaughter through portion control, Stein's is an extension past that point, with coating, breading, and cooking. According to officials at both companies: "Our mutual technical facilities and complementary processing equipment greatly expands our ability to give customers a total package of quality products and service expertise—all from the same corporate source." The combined sales of the two firms after the acquisition were estimated to be in the range of $25 to $30 million.

MR. BUTTNER'S RESPONSE

While Mr. Buttner had conducted several research projects in the recent past, he did not have any formal reports he could send to the professor. Therefore he decided to summarize the general research procedure he employed rather than describe any one research project in detail. After reflecting awhile about the types of projects he most frequently conducted, he wrote the following letter to the professor.

Dear Professor:

The market research I do for Bettcher Industries is basic industrial research and rarely involves a structured format, such as the highly statistically oriented consumer products market research. As you suggest, it is somewhat difficult to provide documented research case histories that would explain a research project from its inception to conclusion.

In brief, the type of research I do typically starts with the question: "Dan, how big is the market for an X type of machine?" No formalized, step-by-step research procedure is drawn up; there is no proposal or cost justification. I start by gathering some statistical data to quantify the market, such as pounds of ham produced yearly, number of firms producing turkey, head of cattle slaughtered yearly. These data give me a feel for the upper limit. I then rely on various contacts in the government, trade associations, trade magazines, and industry to further limit and quantify the market. The information I am looking for includes who uses machines of this type, what they do with them, how many they use.

This step is probably the most important part of industrial market research. Knowing who to call and where to get information usually takes a long time—years. Developing the ability to get the information after you have found the right person is also somewhat difficult. You get a lot of no answers before you develop the right technique. I do most of this search by phone and some of it in person.

After I understand the market limitations, I contact potential end users and ask them specific questions, such as, Would you purchase a machine of this type? How would you use it? Who else would have a need for this machine? How are you doing the job now?

Finally, I summarize these data, including companies and names of people contacted, and present an estimate of market size to management, including all pluses and minuses I have discovered, such as impending technological changes, competition, economic barriers.

Feel free to call me if you think I can be of more help.

Very truly yours,

Daniel A. Buttner
Marketing Research Manager

QUESTIONS

1. Compare the research steps used by Mr. Buttner with the components of the research project we discussed.
2. Would you classify the research process described in Mr. Buttner's letter as exploratory or conclusive? Discuss your answer.

3. Critically evaluate the type of marketing research conducted by Bettcher Industries. What are its pluses and minuses?
4. Could Bettcher benefit from using different research procedures or from conducting research other than the type described in Mr. Buttner's letter? If so, what are they?

CASE 2.2

FAST-TEST: To Use or Not to Use?

Nancy Diamond was recently hired as the marketing manager for the Beauty Aids Division of General Products Corporation. During the past week she has been wondering about the cost-effectiveness of using Fast-Test, a new-product evaluation service. Fast-Test is offered by Innovations Research, Inc., a well-known marketing research firm specializing in evaluating new products and estimating the sales potential for them.

Nancy is particularly interested in the Fast-Test service because the Beauty Aids Division has just developed a new hair spray for which she has complete marketing responsibility. In fact, one of the first challenges of her new job is to decide fairly quickly whether or not the new hair spray should be introduced.

COMPANY BACKGROUND

General Products Corporation is a leading marketer of health care, personal-hygiene, and beauty aid products, with annual sales in excess of $100 million. The company is divided into three divisions, one corresponding to each of its three product groups. The Beauty Aids Division is the largest of the three and accounts for over half of total sales revenue.

Products marketed by the Beauty Aids Division include a variety of hair sprays, hair-coloring products, hair conditioners, shampoos, facial creams, and makeup products. The division has about 30 different brands of these products on the market. Competition for market share is extremely fierce for all the products marketed by the Beauty Aids Division. As a result, new-product introductions, as well as the exit of unsuccessful products, are numerous and frequent in the division's markets. Over the past several years the division has introduced an average of about six new products a year. Fortunately, many of these new-product introductions have fared much better than those introduced by the division's competitors. According to company sources, at least part of the above-average performance of the division's new-product entries can be attributed to the Fast-Test service, which Nancy's predecessor had routinely used to screen all new products prior to market introduction.

FAST-TEST

Innovations Research, Inc., describes its Fast-Test service as a very reliable method of quickly predicting what the market share for a new product will be at the end of the first year after introduction. An attractive feature of this service is that it does not involve any actual test marketing. Thus Fast-Test can predict the likely market performance of a new consumer product without incurring the enormous expense, time delay, and competitive risk associated with a traditional test-marketing operation.

Fast-Test is apparently based on sophisticated data collection and analysis procedures. While technical details about these procedures are confidential, Innovations Research releases a brochure describing the key elements of generating Fast-Test market share predictions for new products. Briefly, once a client firm has developed samples of a new product and drawn up a tentative marketing plan for it, the Fast-Test process consists of the following sequence of steps.

1. Selecting a representative sample of target customers.
2. Exposing the potential customers to promotional materials about the new product.
3. Measuring customers' level of interest in and preference for the new product.
4. Providing customers pricing information on the new product and measuring their intentions to buy the product.
5. "Selling" samples of the new product to customers interested in buying it at a price somewhat lower than the proposed introductory market price.
6. After a few weeks, measuring the reactions of those who bought the product, including the likelihood of their buying the product again.
7. Feeding all the data collected in the previous steps (along with data about market characteristics, consumer demographics, and competing products already on the market) into a computer simulation model to predict year-end market share for the product.

Past users of Fast-Test on Innovations Research's client list include over a hundred consumer product marketers, several of which are *Fortune 500* companies.

FAST-TEST'S TRACK RECORD

The Beauty Aids Division has used Fast-Test to evaluate 20 new products over the past three years. Fifteen of these products were introduced into the market, while the rest were dropped on the basis of Fast-Test's predictions. Specifically, Nancy's predecessor had used a cutoff market share of 4% in deciding whether to introduce a new product. In other words, any new products whose estimated market share at the end of the first year (as predicted by Fast-Test) was below 4% was not introduced.

Upon questioning her colleagues about the rationale for the 4% cutoff share, Nancy discovered two reasons:

1. Top management was very reluctant to continue to support, or even tolerate, new products that did not achieve a "respectable" market share by the end of the first year after introduction.
2. Historically, new products that did not gain at least a 4% market share by the end of the first year had tended to lose rather than make money.

Nancy felt these reasons made sense, especially since she had been told that a key criterion used by top management in conducting annual reviews of the marketing manager's performance was the profitability of new products introduced during the year.

Wondering just how good Fast-Test had been in predicting the market performance of her division's new products, Nancy compiled all the data she could find on products considered for introduction during the past three years. Compiling the data took her a while because she had to go over numerous sales records and related documents. But her effort was well worth it because she was able to uncover the following information:

1. Fast-Test's market share predictions for each of the 20 new products evaluated over the past three years.
2. Actual year-end market shares for 15 of those products that were introduced.
3. Profits (or losses) generated by each of them during their first year on the market.

From an analysis of the data Nancy was able to compute the prediction errors made by Fast-Test. She defined *prediction error* as the difference between the actual and the predicted market shares. The results of her analysis are summarized in Table 1. Prediction errors could not be computed for products 16 through 20 in Table 1 because they were not introduced into the market; Fast-Test had predicted a market share of less than 4% for each of them.

During her investigation Nancy also discovered that products 4, 9, and 11 had been withdrawn from the market a year after introduction.

Table 1 *Comparison of Fast-Test's Predictions with Actual Results*

New-Product Number	Predicted Market Share (%)	Actual Market Share (%)	Prediction Error: Actual Market Share − Predicted Market Share
1	9	8	−1
2	8	8	0
3	5	6	+1
4	6	3	−3
5	7	5	−2
6	8	9	+1
7	4	6	+2
8	5	6	+1
9	5	3	−2
10	6	7	+1
11	4	2	−2
12	5	7	+2
13	7	8	+1
14	6	7	+1
15	5	5	0
16	1	—	—
17	3	—	—
18	0	—	—
19	2	—	—
20	2	—	—
Average*	6	6	0

Note: All percentages have been rounded to nearest integers.
*Average is based only on products 1 through 15.

These three products, in addition to failing to achieve the critical 4% market share, had each lost money. On the average they had resulted in a net loss of about $400,000 apiece. Hence these products were considered to be market failures, although Fast-Test had predicted their market shares to be 4% or better. The remaining 12 products were all considered to be market successes—indeed, all of them were still on the market. Moreover, Nancy's analysis of profitability data showed that the average net profit contribution during their first year on the market was about $500,000.

NANCY'S DILEMMA

Nancy felt she had assembled some interesting data about Fast-Test's predictions. But she felt somewhat frustrated because she did not know what the data meant for the *specific problem she faced,* namely, whether to use Fast-Test for the Beauty Aids Division's newest hair spray. On the one hand, she felt like simply continuing her predecessor's routine practice of using Fast-Test for all new products, especially since she believed the new hair spray would face market conditions similar to those faced by many new products recently considered by her division. On the other hand, she was not sure whether the benefit of using Fast-Test would outweigh the service's hefty price of $80,000. In fact, Innovations Research had just raised the price for conducting a Fast-Test from $75,000 to $80,000.

Prior to joining the Beauty Aids Division of General Products Corporation, Nancy had worked for about five years as an assistant marketing manager for a packaged-goods firm. Her previous employer did not spend much money on marketing research and only infrequently used the services of marketing research firms. During Nancy's association with that firm, she had never seen the marketing manager approve any marketing research project costing over $10,000. Indeed, Nancy believed that the general perception of the firm's top management concerning marketing research was: "Why pay someone to tell you what you already know?" Having gained experience in that kind of environment, Nancy found it particularly difficult to accept Fast-Test.

However, Nancy did not want her previous employer's skepticism to unduly influence her decision making, especially since her intuition—from what she had heard and learned about Fast-Test so far—seemed to indicate that the service was probably worth its price. She was particularly impressed by the computations that showed relatively small prediction errors, which also averaged to an error of 0%. Thus she even wondered whether Fast-Test was not error-free, or perfect, on the average. Could she somehow *verify* her intuition?

In pondering this question, Nancy remembered that in a marketing research course she had been exposed to something called a Bayesian analysis approach to determining the monetary worth of a research project. So she pulled her marketing research textbook off her bookshelf and started going through its table of contents.

QUESTIONS

1. Is Fast-Test "perfect" on the average, as Nancy believes it may be? Why or why not?
2. Can the Bayesian analysis approach be used to evaluate the worth of using Fast-Test to predict market share for the Beauty Aids Division's new hair spray? What assumptions would you have to make in order to use it? How realistic will each of those assumptions be?
3. Should Nancy use Fast-Test for predicting the new hair spray's market share? Develop a qualitative as well as quantitative rationale for your answer.

CASE 2.3

Superior Landscape Company

THE COMPANY

Superior Landscape Company (SLC) is a firm selling a variety of lawn products (e.g., fertilizers, weed killers, privacy fences, decks, sprinklers) as well as landscaping and lawn maintenance services (e.g., landscape design and installation, lawn mowing, fertilizing). The SLC is five years old and is owned by John Norwood, who also holds the title of president. In addition to Mr. Norwood, SLC has four other employees, a salesperson in the retail store operated by SLC, a landscape specialist who designs landscapes and offers landscaping advice to customers, and two servicepeople who provide on-site services such as lawn mowing, fertilizing, and planting trees and shrubs.

THE LOCAL ECONOMY

The SLC is located in a city with a population of 125,000 in the southwestern United States. The local economy is supported by a number of small, but rapidly growing, high-technology firms. The local economy is expected to expand at a fast pace in the coming years. The city's population is expected to grow to 150,000 by the end of the next five years.

SALES AND PROFITS

Mr. Norwood is pleased with the sales growth of his firm. Its total sales in the past year were $335,000, up from $266,000 during the first year the firm opened for business. Additional data about changes in total sales and changes in the contributions made by lawn products and landscaping/lawn maintenance services to total sales are shown in Table 1.

Table 1 *Summary of SLC's Sales*

Year	Sales of Lawn Products	Sales of Landscaping and Lawn Maintenance Services	Total Sales
1	$213,000	$53,000	$266,000
2	$212,000	$60,000	$272,000
3	$220,000	$73,000	$293,000
4	$228,000	$88,000	$316,000
5	$235,000	$100,000	$335,000

Mr. Norwood is also pleased with the profit performance of SLC. Profits before taxes, as a percentage of sales, have averaged about 20% during the past five years. During the past year SLC's profits before taxes were about $75,000.

COMPETITOR'S MARKETING RESEARCH STUDY

Mr. Norwood recently discovered that a major competitor of SLC had just completed a personal-interview survey of 200 customers in the area. The survey was conducted by a local marketing research firm. Mr. Norwood estimated that the survey cost the competitor about $4000.

Mr. Norwood did not know the details of the survey—particularly, what the findings of the survey were. However, one of SLC's customers who had participated in the survey volunteered some information about the personal interview. The interview lasted about 30 minutes and dealt with a variety of issues, such as the types of lawn-related products and services used by the respondent, annual expenditures on lawn care, and the respondent's perceptions about various lawn care/landscaping firms in the area.

MR. NORWOOD'S DECISION

The idea of conducting a consumer survey intrigued Mr. Norwood. He believed that a survey similar to the one conducted by SLC's competitor could provide valuable information about customer needs and preferences regarding lawn-related products and services. He also felt that such a survey could uncover customer perceptions about SLC's products and services versus those offered by key competitors. In Mr. Norwood's opinion information from a survey of this nature would be of great help in formulating SLC's future marketing strategy and strengthening its market position in the coming years. He therefore made a note to call the local marketing research firm and make plans for conducting a survey similar to the one conducted by SLC's competitor.

QUESTIONS

1. Should Mr. Norwood commission a survey similar to the one conducted by SLC's competitor? Why or why not?
2. What kinds of information are likely to be most helpful in planning SLC's future marketing strategy? How should such information be obtained? Justify your answer with specific reasons.
3. Is there any ethical issue involved in Mr. Norwood contacting the marketing research firm that conducted the survey for SLC's competitor? Should the marketing research firm agree to conduct a survey for SLC if Mr. Norwood requests one?

PART THREE

Data Collection PROCEDURES

CHAPTER 5 *Secondary Data*

CHAPTER 6 *Overview of Primary-Data Collection Methods*

CHAPTER 7 *Qualitative Research*

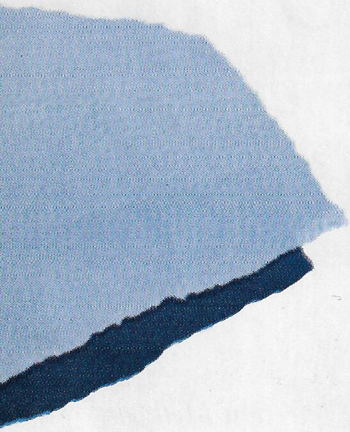
CHAPTER 8 *Introduction to Experimentation in Marketing*

CHAPTER 9 *Selected Experimental Designs*

CASES FOR PART THREE

CHAPTER 5

Secondary Data



CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *State the three advantages of secondary data over primary data.*
- *Describe the secondary-data features that one should examine to assess the relevance and the accuracy of the data.*
- *Distinguish between (1) original and secondhand sources of secondary data and (2) internal and external sources of secondary data.*
- *Summarize the relative advantages and drawbacks of secondary data obtained from government sources over secondary data obtained from syndicated sources.*
- *Define secondary-data management and discuss why it is becoming increasingly important.*
- *Given a description of the data needed in a research project, identify appropriate secondary-data sources and obtain the necessary data from them.*

Secondary Data
Data collected for some purpose other than the research situation.

The notion of secondary data was introduced in previous chapters. ***Secondary data*** are data collected for some purpose other than the research situation at hand. They are usually readily available. The following are all examples of secondary data.

- Data on warranty cards (e.g., date, place, and purpose of purchase) mailed to an appliance manufacturer by recent buyers of the firm's appliances.
- Nielsen television ratings indicating the relative popularity of various TV programs in a certain week.
- Unemployment rates in different sections of the country, estimated and released by the U.S. Department of Labor.
- Data made available by an automobile dealers' association to its members regarding the volume and models of used cars sold in a certain part of the country during a given period.
- Data reported by Ronald Milliman in an article published in the *Journal of Marketing* (a professional journal of the American Marketing Association) indicating a relationship between the tempo of background music played in a supermarket and the amount of purchases made by supermarket customers.[1]

As the examples illustrate, a wide variety of secondary data are available for possible use by a researcher. However, the purpose for which secondary data were originally gathered will normally differ from the objective of the specific project the researcher is contemplating. After all, secondary data, by definition, are data collected for purposes other than specific research objectives. Thus a marketing researcher faces the twofold challenge of identifying the various secondary data that may be useful and evaluating such data vis-à-vis the requirements of the research study. The purpose of this chapter is to present details concerning this challenge and to offer guidelines for dealing with it.

---

[1] Ronald E. Milliman, "Using Background Music to Affect the Behavior of Supermarket Shoppers," *Journal of Marketing*, 46 (Summer 1982), pp. 86–91.

Let us first examine the pros and cons of secondary data in order to highlight why a researcher must seek out such data and, at the same time, exercise caution in using them.

ADVANTAGES OF SECONDARY DATA

Cost and Time

Primary Data
Data collected specifically for a project.

Secondary data obviously have cost and time advantages over ***primary data,*** which are data collected specifically for a project. A research study will invariably be less expensive and easier to complete if it utilizes secondary data rather than primary data. As an illustration, consider the following hypothetical situation:

> **Example.** The American Lung Association (ALA) has just developed an antismoking campaign, consisting of television and radio commercials warning the public about the hazards of cigarette smoking. Although the ALA would like to launch the campaign nationally, it faces a serious budget constraint at the present time. However, rather than delay the campaign launch until its financial position improves, the association wants to introduce the campaign selectively in states where it is likely to do the most good. For this purpose, the association must identify states with relatively high per capita consumption of cigarettes. How can the ALA identify such states?

One option available to the ALA is to conduct surveys of, say, 1000 people in each state to collect the data needed to estimate per capita consumption. As you might imagine, such a primary-data collection approach will be expensive and time-consuming. For instance, even if the ALA spends only about $500 per state for collecting and analyzing data—which is a conservative cost estimate—it will have to spend a total of $25,000. Given its tight financial situation, the ALA would be unwise to spend that kind of money on primary-data collection, especially since useful secondary data are likely to be available at much lower cost.

Several potentially productive secondary-data collection approaches are available for estimating per capita cigarette consumption on a state-by-state basis. For instance, data on state cigarette tax revenues (available through published state government documents) can be used to estimate total consumption of cigarettes in a state. Per capita cigarette consumption in the state can then be determined by dividing total consumption into the number of people in the state (available through U.S. Census

Bureau data). Alternatively, perhaps someone else—such as an academic researcher, a commercial marketing research firm, a governmental agency, or a trade association—has already collected data about cigarette consumption in various states. The availability of such data can be ascertained by a trip or a telephone call to a library.

Invariably, plunging into primary-data collection without first looking into the availability of secondary data is a cardinal mistake. Researchers making this mistake will be squandering valuable resources—money and time. Using secondary data to ascertain such things as demographic compositions of markets, retail sales of a product category, changes in brand shares over time, and market potentials for industrial products can be particularly cost- and time-efficient because a variety of rich secondary-data sources are available for these purposes.

Availability

In addition to their obvious cost and time benefits over primary data, secondary data have a more subtle advantage. Certain kinds of data a firm may find difficult to collect on its own as primary data may be available as secondary data. For example, the Census of Retail Trade (published by the U.S. Census Bureau) contains data on the sales of retail establishments, classified according to numerous categories of outlets. This census may be the only source of accurate data available to a manufacturing firm selling its products to retailers and wishing to ascertain the size of its markets as measured by the sales of different categories of retail outlets. Given the sensitive nature of the data, the manufacturing firm may find that gathering accurate and comprehensive data on its own is difficult. However, the U.S. Census Bureau, by virtue of its legal authority, can collect and disseminate such data. Another federal agency that has similar authority is the Securities and Exchange Commission (SEC). A variety of publicly held business firms in the United States are required by law to file certain financial data with the SEC, which, in turn, makes them public.

While some forms of secondary data are available free (e.g., data contained in most government publications and in research studies published in books and journals), those made available by commercial marketing research firms are not. Secondary data sold by research firms are usually called ***syndicated data,*** or ***syndicated services.*** The cost of collecting syndicated data is spread over a number of client firms; the cost to any one firm is invariably lower than the cost the firm would incur were it to collect the data on its own. Indeed, firms with tight budgets and other resource constraints simply cannot afford to gather primary data about their markets and the performance of their products in those markets.

Syndicated Data (Services) *Secondary data generated by research suppliers on a regular basis and sold to many clients.*

As an illustration, consider a small firm manufacturing a line of canned vegetables and distributing its brands to supermarkets across the country through food brokers. The firm has no direct contact with the supermarkets or the customers who buy its brands. Therefore it may be unable to learn much about who buys its brands, what its brands' shares are in various regional markets, and so on, if syndicated services—such as the National Purchase Diary Panel and the Nielsen ScanTrack Services[2]—are not available.

The point of the preceding discussion is to emphasize an important, although somewhat subtle, benefit of secondary data: Such data, especially those available through government documents and syndicated services, can shed light on certain aspects of markets, products, consumers, and so on, that may be extremely difficult, if not impossible, for a firm to study on its own.

LIMITATIONS OF SECONDARY DATA

Relevance

Secondary data are not without drawbacks. One major drawback relates to the relevance of the data for the specific research purpose. Available secondary data may not match the data needs of a given project on one or more of the following factors: (1) the *units* in which the data are measured, (2) the *category breakdowns or definitions* of variables for which the data are reported, and (3) the *time period* during which the data are measured.[3]

Data on many variables of interest can be gathered by using a number of alternative units of measurement. Sales of a product can be measured in dollars or in physical units; product shipments can be measured in terms of volume, weight, dollar value, or truckloads; a consumer's education level can be measured in terms of the highest degree held or the number of years of formal education completed. A particular choice of measurement units by the source of secondary data may render the data useless for certain research purposes, as the following example illustrates.

Example. Carpets Unlimited and Sentinel Corporation are two newly started firms. Carpets Unlimited manufactures a variety of carpets, and Sentinel Corporation produces a line of smoke detectors. Both firms are eager to estimate the total residential-market potential for their products in different sections of the country.

One surrogate estimate of potential for their products in a region is the total size of residential-housing units in that region. However, size of residential housing can be measured in at least two ways: number of rooms and square feet of area. Residential-housing size expressed in number of rooms may be adequate and appropriate for the data needs of Sentinel Corporation but not for the data needs of Carpets Unlimited. A meaningful estimate of market potential for carpeting requires size data expressed in square feet.

---

[2] The data made available through these syndicated services are described in the following brochures: "Sharpening Marketing Decisions with Diary Panels," prepared (in1975) by NPD Research, 900 West Shore Road, Port Washington, N.Y.; 11050 and "Nielsen ScanTrack Services" prepared (in 1986) by the A. C. Nielsen Company, Nielsen Plaza, Northbrook, Ill. 60062.

[3] These three factors are outlined in a number of marketing research textbooks; see, for example, Harper W. Boyd, Ralph Westfall, and Stanley F. Stasch, *Marketing Research—Text and Cases* (Homewood, Ill.: Irwin, 1985), p. 162.

The U.S. Census of Population and Housing contains a variety of data on residential housing that may be quite useful to Sentinel Corporation in estimating market potential for smoke detectors. Data are available on total number of rooms, number of bedrooms, year in which a unit was built, and number of stories in the unit. However, these data do not yield meaningful estimates of market potential for residential carpeting.

Thus, in general, while data on a variable of interest (e.g., size of residential housing) may be readily available, the relevance of data for a particular application depends on the measurement units used. However, one must not hastily dismiss secondary data as being irrelevant just because the measurement units do not exactly match the requirements of a project. In certain situations converting the reported data into the specific units of interest may be possible. For example, a trucking company interested in the *volume* of industrial chemicals transported may find secondary data reported in *weight* units. But the secondary data here are useful since they can be easily modified to meet the trucking company's needs by using appropriate weight-to-volume conversion factors. Similarly, total *unit* sales of some product, say television sets, can be derived from total *dollar* sales data if a meaningful estimate of the average price per television set is available. Of course, such conversions may not always be possible. Data on total television set sales reported in dollars will be useless if one is interested in unit sales of different sizes of television sets rather than total unit sales.

Secondary data may also be irrelevant if the categories for which data are summarized are not consistent with the categories of interest to a researcher. Consider a marketer of products and services primarily aimed at senior citizens. This marketer is interested in obtaining data about the life-styles and activities of different age-groups of senior citizens. Suppose a comprehensive study of life-styles and activities of adults classified by different age-groups is readily available. The life-style and activity data are summarized in the study according to the following five age-groups:

18 to 25 26 to 35 36 to 45 46 to 55 over 55

This study is not useful to the marketer because even though the study contains data on relevant variables, the categories chosen for the age variable are not consistent with the marketer's specific needs. The highest age-group for which data are summarized in the study is "over 55"—a category too broad to be relevant to the marketer.

The mismatch between categories as defined by a secondary-data source and as desired by a decision maker may not always be as obvious as in this illustration. The difference may be quite subtle and, if a secondary-data user fails to recognize it, can lead to erroneous inferences. This point is forcefully made by a situation reported in *American Demographics* magazine.

> Are metropolitan areas markets? Consider the case of Brownsville, Texas. A chain store whose products are geared to an upscale clientele argued that Brownsville deserved a high priority for a new store because retail sales per household there were much higher than the national average.
>
> Brownsville is a border town; and the standard metropolitan statistical area [SMSA] of which it is a part, Brownsville–Harlingen–San Benito, consists of Cameron County. According to the 1980 census, the SMSA's median household income is only $11,700—well below the $16,800 national average.
>
> With income so low, how could retail sales be so high? The answer is that Mexicans routinely cross the border to shop. Thus, dividing total retail sales by the number of county households inflates the real spending attributed to the average Brownsville household. For this chain to open a new store in the area would be a mistake, because few upscale residents live there.[4]

The inflated per-household retail sales figure resulted from the chain store's failure to recognize the distinction between the market, or *trading area,* of retail outlets in Brownsville and the Census Bureau's definition of the area—namely, Cameron County—for which data on retail sales and number of households were readily available.

Researchers using census data for studying the nature of geographic markets must be especially cautious. Well-defined geographic area (e.g., counties or states) for which census data are summarized typically do not coincide with trading areas surrounding core cities or SMSAs. From an analysis of secondary data on variables most likely to be associated with consumer's shopping behavior—such as place of residence and work, commuting patterns, and newspapers read—the U.S. Bureau of Economic Affairs has divided the United States into 183 economic areas, whose boundaries are very different from traditional SMSA, county, or state boundaries.[5]

---

[4] Thomas I. Rubel, "Metros, Markets . . . and More," *American Demographics,* © July 1983, p. 22, excerpted with permission from *American Demographics,* Ithaca, N.Y.

[5] For a discussion of the procedure used in creating the economic areas, see Rubel, 'Metros, Markets . . . and More," pp. 22–25, 44. The possible lack of relevance of census division boundaries for marketing applications is further supported by a theory proposed by Joel Garreau in his book *The Nine Nations of North America* (Boston: Houghton Mifflin, 1981). From economic, social, cultural, political, topographical, and natural-resource factors, Garreau claims that North America can be divided into nine distinct "nations."

Table 5.1 *Changes in Variable Definitions Between the 1970 and 1980 Censuses*

Farm definition	Old definition can be reconstructed
Householder	Replaces household head concept
Industry	Many code changes; classification system changed somewhat
Occupation	Classification system substantially changed
Poverty definition	Some adjustments, but old definition can be reconstructed
Race	Several categories added; revised coding rules affect "White" and "Other"
Residence in 1975/place of work	Broad categories in 1970, but coded to county group in 1980
Telephone in unit	Narrower than telephone availability in 1970
Work disability	Not counted if it has lasted less than six months

Source: Adapted from Paul T. Zeisset, *Public-Use Microdata Samples from the 1980 Census,* U.S. Department of Commerce, Bureau of the Census, Data User Services Division (Washington, D.C.: U.S. Government Printing Office, 1983), p. 9.

Another facet of the category problem of which secondary-data users must be aware relates to *changes* in category definition that may occur from one time period to the next. For instance, in the 1970 census Hispanics were classified as whites, while in the 1980 census they were assigned a separate race category. Before drawing inferences from secondary data obtained from multiple time periods, one must ensure that no changes occurred in the definitions of variables or their categories. Table 5.1 summarizes the definition changes for certain variables between the 1970 and 1980 censuses of population. Obviously, researchers interested in studying the changes in variables between 1970 and 1980 must first determine whether they can meaningfully compare the data over time.[6]

The time period during which secondary data were collected is the third factor affecting the data's relevance for a specific application. Using data that are too old can be more dangerous than using no data at all, especially in making decisions relating to rapidly changing or highly volatile markets. Consider, for instance, the market for personal computers. Given the state of turmoil in this market—with rapidly changing competitors, computer models, prices, and applications—secondary data on, say, the demographics of personal-computer owners will most likely become useless shortly after they are collected.

---

[6] An interesting discussion of faulty inferences stemming from failure to recognize definition changes is given in "10 'Shortcuts' to Evaluating Demographics Statistics," *Marketing News,* September 17, 1982, Section 1, pp. 9–10.

Data published by the U.S. Census Bureau are particularly prone to being out of date. To illustrate, the Census of Population and Housing is conducted only once every 10 years. Moreover, there is a lag of about two years between collection and publication of the data. Fortunately, difficulties stemming from the recency of secondary data are somewhat mitigated by two factors: Secondary-data sources usually provide periodic updates for their data, and rapid advances in computer and communication technology continue to reduce the lag between data collection and dissemination. The U.S. Census Bureau, for example, issues *Current Population Reports,* which update population data on an annual basis for major market areas. Similarly, some marketing research firms as well as state and local agencies periodically update census data on the basis of their own research.[7] Certain syndicated services (e.g., Chase Econometrics and DRI, Inc.) allow clients' computers direct access to their own computer files containing the most current demographic data.[8]

In summary, the relevance of secondary data depends on three factors: (1) measurement units; (2) variable and category definitions; and (3) data recency. A mismatch between any of these and the specific needs of a research study will reduce the data's relevance. Under such circumstances, however, a researcher must not be hasty in concluding that the data are useless. The researcher should explore whether the data can be modified in some way—for example, converted to different units, recast into different categories, or updated—so as to render them useful. While making such modifications (assuming they are feasible) will require some effort, it is still likely to be cheaper and less time-consuming than collecting primary data.

Accuracy

Secondary-data users, in addition to ensuring the relevance of such data, must verify their accuracy. The concepts of relevance and accuracy do overlap somewhat because they jointly determine the potential usefulness of a particular set of secondary data. However, the two terms refer to distinct facets of usefulness. While relevance focuses on the *suitability* of the data (i.e., the form and recency of the data), accuracy focuses on the *trustworthiness* of the data. Obviously, the usefulness of secondary data will be diminished if they lack *either* of these two traits.

---

[7] See William J. Klocke, "Don't Overlook Secondary Data When Researching Retail Markets," *Marketing News,* May 14, 1982, Section 1, p. 17.

[8] See Doris L. Walsh, "Giving Demographics More Byte," *American Demographics,* April 1983, pp. 18–23, 47–48.

The key to assessing the accuracy of secondary data lies in learning as much as possible about the process involved in collecting the data. This task may not always be easy and may require some effort on the part of researchers planning to use secondary data. Indeed, in many cases a major limitation of secondary data is the lack of information necessary to evaluate the data's accuracy rather than their degree of accuracy per se. After all, a researcher's decision about whether to use a particular set of secondary data will be relatively simple *if* adequate information is available to ascertain the data's trustworthiness. Unfortunately, the proliferation of easily accessible secondary data, on the one hand, and the scarcity of readily available information to evaluate them, on the other, may tempt some researchers to downplay the issue of accuracy. This temptation should be avoided; with a little effort researchers can usually get at least some feel for the trustworthiness of secondary data.

The following set of interrelated questions plays a cenral role in evaluating secondary-data accuracy: *Who* collected the data? *Why* were the data collected? *How* were the data collected? Answers to the who, why, and how questions can provide valuable insights about the trustworthiness of the data. Let us now examine each of these questions.

Who Collected the Data? The reputation of the sources of secondary data is an indicator of data accuracy. Well-known governmental sources like the U.S. Bureau of Census and many commercial marketing research firms, especially those established in the profession for a long time, have a reputation for setting and maintaining high data quality standards. Secondary data from such sources require far less scrutiny than data offered by organizations not well-established within the research profession or by sources whose credibility and competence are unknown.

Why Were the Data Collected? An examination of the explicit, or even implicit, reason for the data collection can sometimes indicate the accuracy of the data. To illustrate, suppose a firm publishing a magazine conducts, on its own, a survey of its readers. The survey gathers data on a variety of factors, such as the demographic profiles of the readers and the impact on them of advertisements appearing in the magazine. The purpose of the survey is to generate readership data helpful to potential advertisers in deciding whether to place advertisements in the magazines. Of course, the publisher also hopes that the data will convince many firms to buy advertising space in the magazine.

Now think of yourself as one of the potential advertisers. Assume you have access not only to the secondary data offered by the publisher but also to similar data offered by a marketing research firm specializing in collecting and selling syndicated data about readers of a number of magazines. Which of these two sources—the publishing firm or the marketing research firm—would you trust more?

You are most likely to trust the marketing research firm more, for two good reasons:

1. Being an independent source, with no hidden bias, the marketing research firm is more likely to provide unbiased data.
2. Since the marketing research firm *specializes* in magazine readership surveys, it is more likely to be capable of generating good-quality data. Notice that this reason also relates to the who question discussed earlier.

These factors do not mean you should not use the data provided by the publishing firm—especially since you would have to pay for the research firm's data, while the data provided by the publisher are free. However, they do imply that before making any advertising decision on the basis of the publisher's data, you should carefully review the process involved in collecting the data. We next turn to the role of the data collection methodology in evaluating the accuracy of secondary data.

How Were the Data Collected? Perhaps the most effective way to judge the accuracy of secondary data is to scrutinize the data collection process, *beyond* the who and why aspects already discussed. What specific data collection instruments were used? How large was the sample? What types of sample units provided the data? During what specific time periods (e.g., time of day, day of week) were the data collected? Answers to questions like these are crucial in assessing data accuracy.

Secondary-data suppliers with good reputations generally publish at least a brief description of their data collection procedures. Most of them will also furnish additional details if desired by potential data users. For instance, Yankelovich Clancy Shulman, a well-known marketing research firm, offers syndicated services including a brochure entitled "The Yankelovich Monitor." The "Monitor" provides annual statistics on numerous social trends and their association with buying behavior within the general population as well as specific demographic segments.[9] Exhibit 5.1 contains a published summary of the procedures used to collect data for the " Monitor." In the second paragraph of the exhibit, notice the research firm's offer to provide additional details about its methodology.

Good documentation of methodology, including statements about limitations of the data, is itself an indicator of high data quality. A secondary-data source that readily offers such documentation and is willing

---

[9] Additional information about this syndicated service is available in "The Yankelovich Monitor," a brochure prepared by Yankelovich Clancy Shulman, 8 Wright Street, Westport, CT 06880.

Exhibit 5.1 *Illustration of a Data Collection Procedure Summary*

HOW ARE MONITOR DATA COLLECTED?

An Annual Survey

The social trend and manifestation data covered by Monitor are obtained by means of a survey conducted at the same time each year among a nationally projectable sample of 2500 adults, 16 years of age and older, including a special subsample of 300 college students living on campus. The sample is based on the most recent Census data. All interviews are conducted in person at the respondent's home, and take approximately $2\frac{1}{2}$ hours to complete.

The size of social trends are determined by compositing scores on item batteries that have undergone extensive pretest and validation procedures. (A comprehensive technical description of Monitor, covering all aspects of its methodology, is available on request. Inspection of the questionnaire employed may also be arranged.)

Ongoing Qualitative Research

New trends and new manifestations are identified by a continuing program of in-depth research conducted among the general public and "precursor" groups of consumers. This program of original research is supplemented by systematic analyses of publications and custom studies conducted by Yankelovich Clancy Shulman on a wide variety of business and social problems.

Source: "The Yankelovich Monitor," prepared by Yankelovich Clancy Shulman, 8 Wright Street, Westport, CT 06880.

to submit its procedures for additional examination will generally have high quality standards. By the same token, organizations unable or unwilling to document their data collection methodologies are suspect. A researcher must be especially cautious in dealing with secondary data coming from such organizations.

Original Source *A term used in the context of secondary data; it refers to the source that actually collects the data.*

Secondhand Source *A term used in the context of secondary data; it refers to a source that uses data collected by an original source to generate its own summaries, interpretations, and the like.*

One's ability to evaluate the accuracy of secondary data also depends on whether the data are from the original source or a secondhand source. The ***original source*** is one that actually collects the data. A ***secondhand source*** is one that uses data collected by the original source to generate its own summaries, interpretations, and the like. For example, a report issued by the Census Bureau on data gathered through the Census of Population and Housing is the original source of such data. However, a report on population characteristics issued by an independent organization, which used data from the Census of Population and Housing as the basis for its report, is a secondhand source of population data.

The original source of secondary data is more likely than a secondhand source to spell out the procedures used in generating the data.

Moreover, a secondhand source may unintentionally, or in some cases even deliberately, restate or reinterpret the original source's data in inappropriate ways. A secondhand source may also make mistakes in transcribing the data from the original source and thus provide incorrect numbers. Hence, before using secondary data available through a secondhand source, one must identify the original source and examine the procedures used, if possible. As a general rule, one should obtain all the relevant data from the original source, if at all feasible.

SOURCES AND TYPES OF SECONDARY DATA

Internal Source *A source within the organization.*

External Source *A source outside the organization.*

There are literally thousands of potential sources of secondary data—too numerous to list in a book of this nature. However, from the point of view of a decision maker's organization, a useful technique is to group secondary-data sources broadly into internal sources and external sources. As the terms imply, ***internal sources*** are sources within the organization, while ***external sources*** are sources outside the organization. Secondary data are typically labeled internal secondary data or external secondary data, depending on their source.

Internal Secondary Data

A firm's historical record of sales, a public service association's list of donors, past records of services rendered to patients by a hospital, and public opinion polls conducted in the past by a political candidate's campaign office are all sources of internal secondary data for the respective organizations. We earlier emphasized the usefulness of looking into the availability of secondary data before plunging into primary-data collection. Likewise, when one is seeking secondary data, checking out the availability of internal secondary data before turning to external sources is beneficial. Internal secondary data, if available, may be obtained with less time, effort, and money than external secondary data. In addition, they may also be more pertinent to the situation at hand since they are from *within* the organization. These points are illustrated in the following example.

Example. The publisher of a Houston newspaper wants to know the "penetration" of the newspaper in the various communities it serves. Specifically, the publisher would like to determine the proportion of households subscribing to the newspaper in each community. A marketing research firm specializing in newspaper audience research has estimates of household penetration for various newspapers in the Houston metropolitan area. These estimates are based on a survey of a representative sample of households in the Houston metro area and are updated every six months. Should the publisher purchase this syndicated service from the marketing research firm?

The penetration estimates provided by the syndicated service, an external secondary-data source, will certainly offer useful insights to the publisher. However, the addresses of current subscribers—internal secondary data that the publisher should have—will be even more beneficial. While the syndicated service is available only for the Houston metropolitan area, addresses of subscribers can pinpoint relative concentrations of subscribers in *all* areas served by the newspaper. These concentration estimates, coupled with readily available census data on total number of households in the corresponding areas, can provide the penetration estimates the publisher desires. Thus the total spread of the geographic area covered by the internal secondary data should be more in line with the publisher's needs. Furthermore, although the syndicated service is updated every six months, the newspaper's subscriber list will be up to date all the time. In short, given the specific objective of the publisher, internal secondary data will be more relevant and less expensive than the syndicated data offered by the research firm.

Of course, internal secondary data may not always be suitable or adequate. For instance, if data are needed on the relative geographic penetrations of the publisher's newspaper and those of competing newspapers, the publisher may have to buy the syndicated data offered by the research firm. In fact, the publisher may even have to collect primary data if it is keen on knowing the competitive penetration rates in communities outside the Houston metro area. But the point of this discussion is that once data needs are identified, a systematic search for data will pay off handsomely. It will ensure the gathering of suitable data without undue expenditures of time, effort, and money.

Exhibit 5.2 shows a flow diagram outlining the sequence of steps in a systematic data search. The logical starting point for any data search should be an examination of internal secondary-data sources. If the data available within the organization are unsuitable or inadequate, the search should be extended to external secondary-data sources. Primary data should be collected only if absolutely essential. Indeed, the rigor with which the relevance and the accuracy criteria are applied in assessing the adequacy of internal and/or external secondary data should be tempered by available resources (time and money) and the purpose of the research. For instance, if resources are tight and the research is basically exploratory, the researcher can, and perhaps should, be somewhat lenient in evaluating available secondary data.

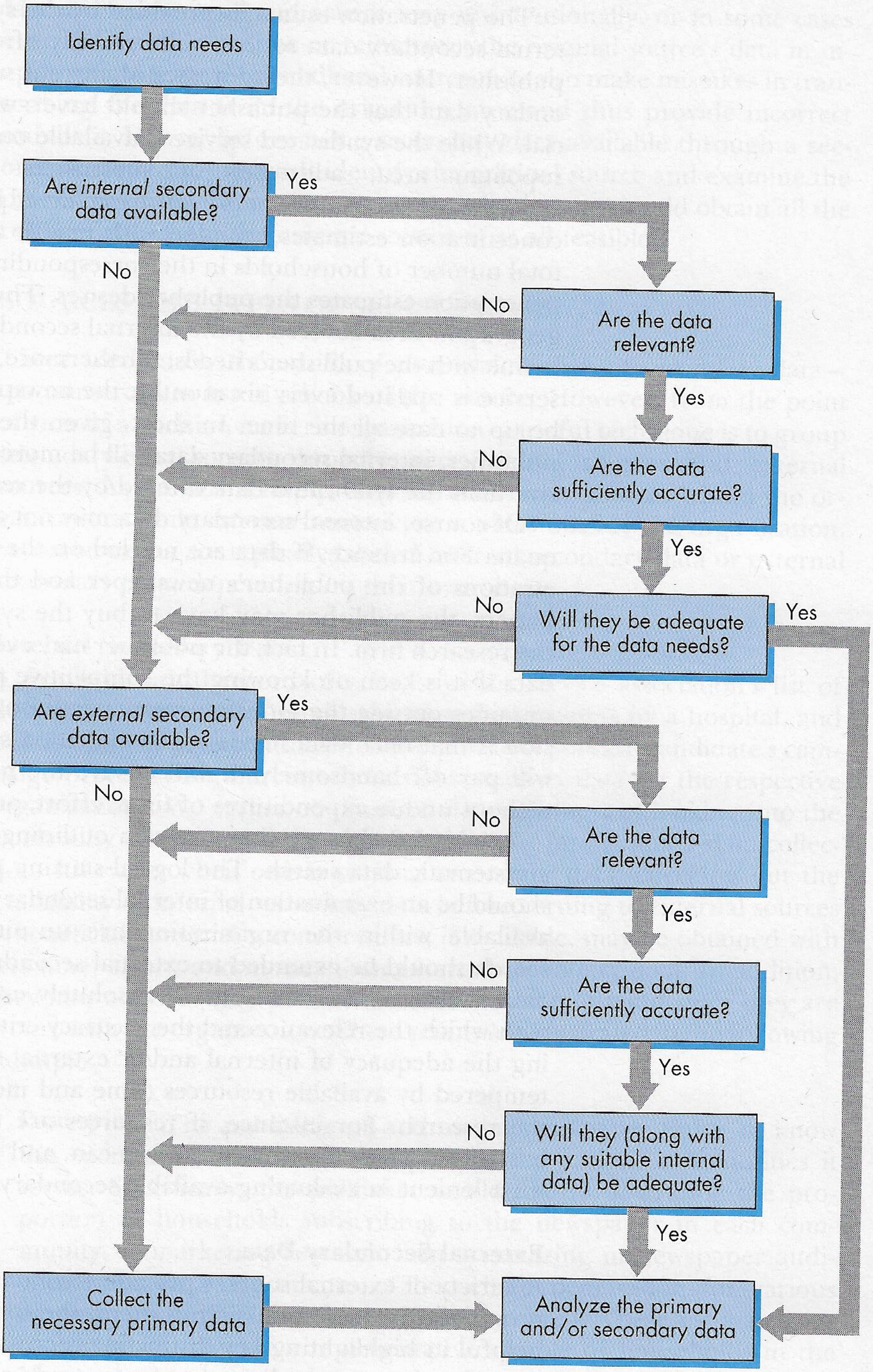
External Secondary Data

A variety of external sources provide secondary data. Although there is no standard way for classifying them, the following groupings may be helpful in highlighting key distinctions across the various sources: government sources; syndicated sources; trade associations; miscellaneous sources; and abstracts, directories, and indexes. Examples of key sources under each of these categories, along with brief descriptions of what

sekalaiset

yht. vedot luettelot hakemistot

Exhibit 5.2 *Flow Diagram for Conducting a Data Search*


Identify data needs
Are *internal* secondary data available?
Yes
No
Are the data relevant?
No
Yes
Are the data sufficiently accurate?
No
Yes
Will they be adequate for the data needs?
No
Yes
Are *external* secondary data available?
Yes
No
Are the data relevant?
No
Yes
Are the data sufficiently accurate?
No
Yes
Will they (along with any suitable internal data) be adequate?
No
Yes
Collect the necessary primary data
Analyze the primary and/or secondary data

they offer, are provided in the appendix at the end of this chapter. Let us now examine the main features of the five broad categories of external sources.

Government Sources. Government agencies at federal, state, and local levels collect more data about people, firms, markets, and so on, than any other secondary-data source. They therefore form an extremely rich reservoir of data for researchers. Moreover, much of these data are free. Looking into government sources of secondary data is a must if the information requirements of a project appear to relate to data in the public domain. Depending on the objective of a project, contacting an appropriate government agency or even paying a visit to the government documents section of a library can turn up potentially valuable published data.

Documents published by government sources are typically in the form of summary reports and tabulations based on the raw data collected. However, if additional details or different tabulations are needed, one may be able to obtain some or all of the raw data, usually for a fee. The U.S. Census Bureau, for example, sells computer tapes containing certain types of raw data collected through the variety of censuses it conducts. A case in point is Public-Use Microdata Files, containing data from the Census of Population and Housing.

Public-Use Microdata Files are raw data records, stored on computer tape, for a representative sample of housing units in the United States. These records provide data on the characteristics of each unit and the people living in it. However, to protect the confidentiality of respondents, the Census Bureau excludes all identifying information from the records. For the same reason microdata records do not identify geographic areas with less than 100,000 people. When census data users have unique needs not met by published documents, microdata files offer them the flexibility to analyze the raw data in whatever manner they desire. Exhibit 5.3 outlines some key capabilities and limitations of microdata files, as described by the Census Bureau.[10]

Syndicated Sources. Syndicated sources consist of marketing research firms offering syndicated services. The well-known Nielsen Retail Index is an example of a syndicated secondary-data source. Brief descriptions of several syndicated services are provided in the appendix at the end of the chapter.

Secondary data offered by marketing research firms are not free. However, since such data are syndicated, their cost is spread over a number of client organizations. Therefore if suitable syndicated data are available, they should be more cost-effective than primary data for any

---

[10] Paul T. Zeisset, *Public-Use Microdata Samples from the 1980 Census,* U.S. Department of Commerce, Bureau of the Census, Data User Services Division (Washington, D.C.: U.S. Government Printing Office, 1983).

Exhibit 5.3 *Features of Public-Use Microdata Files*

USES OF MICRODATA FILES

Public-use microdata files essentially make possible "do-it-yourself" special tabulations. The 1980 files furnish almost the full richness of detail recorded on long-form questionnaires in the census. Subject to the limitations on sample size and geographic identification, it is possible for the user to construct a seemingly infinite variety of tabulations interrelating any desired set of variables. Users have the same freedom to manipulate the data that they would have if they had collected the data in their own sample survey, yet these files offer the precision of census data collection techniques and sample sizes larger than would be feasible in most independent sample surveys.

Microdata samples will be useful to users (1) who are doing research that does not require the identification of specific small geographic areas or detailed cross tabulations for small populations, and (2) who have access to programming and computer time needed to process the samples. Microdata users frequently study relationships among census variables not shown in existing census tabulations, or concentrate on the characteristics of certain specially defined populations, such as unemployed homeowners or families with four or more children.

SUBJECT CONTENT

With only minor exceptions, microdata files contain the full range of population and housing information collected in the 1980 census: 503 occupation categories, age by single years up to 90, income by $10 intervals up to $75,000, and so forth. Because the samples provide data for all persons living in a sampled household, users can study how characteristics of household members are interrelated (for example, income and educational attainment of husbands and wives).

PROTECTING CONFIDENTIAL INFORMATION

Records on public-use microdata samples contain no names or addresses. Also, the Bureau limits the detail on place of residence, place of work, high incomes, and selected other items to further protect the confidentiality of the records. Microdata records identify no geographic area with fewer than 100,000 inhabitants. Microdata samples include only a small fraction of the population, drastically limiting the chance that the record of a given individual is even contained in a microdata file, much less identifiable.

Source: Paul T. Zeisset, *Public-Use Microdata Samples from the 1980 Census,* U.S. Department of Commerce, Bureau of the Census, Data User Services Division. (Washington, D.C.: U.S. Government Printing Office, 1983), p. 3–4.

individual client. In contrast to secondary data from government sources, syndicated data are likely to directly focus on the needs of decision makers. That is, while government sources cater to a wide variety of constituencies, research firms have a more limited and better-defined clientele. Moreover, syndicated data are typically updated more frequently than government data. Some syndicated data are updated as often as once a week.

In short, while data from government sources are generally less expensive than syndicated data, the latter are more likely to be directly relevant to the needs of decision makers. Of course, the two types of data may not necessarily be substitutes for each other; certain kinds of data may only be available through one type of source. To illustrate, detailed and rich data about the general characteristics of people and markets may only be available through the various government census studies; likewise, specific data on market shares or brand shares may only be available through syndicated sources. Thus a choice between government and syndicated sources may sometimes be influenced more by data availability than by cost-effectiveness or relevance considerations.

Although syndicated sources typically sell the same data to all subscribing clients, some permit a certain degree of customization. Specifically, client firms can submit questions of their own for inclusion in the questionnaire routinely used to gather syndicated data. There is an additional charge for including these custom questions. However, the data generated by these questions, and any analysis of such data, will be revealed only to the firms submitting the questions. Thus, to a certain extent, customization of syndicated services offers benefits of both secondary and primary data.

Roper Reports, prepared by the Roper Organization, is one example of a syndicated service offering customization. This service continuously monitors and reports public opinion and behavior concerning a broad range of social, economic, and political issues, as well as various kinds of goods, services, and life-styles. Data are gathered every five weeks through personal interviews with a national sample of 2000 adults 18 years of age or older. In addition to the standard or regularly asked questions, a tack-on custom-question service is available to *Roper Reports* subscribers and nonsubscribers. The cost of the custom-question service depends on question complexity, the number of questions used in a given survey, the number of surveys in which the same questions are repeated, and whether the firm submitting questions is a *Roper Reports* subscriber. According to the Roper Organization: "This service offers a unique combination of frequency, speed of report delivery, quality, low cost, large sample size and extensive demographic breaks."[11]

One interesting syndicated service that is "fully customized" is AIM (abbreviation of Association-Identification-Measure), offered by R. H. Bruskin Associates.[12] The AIM service is conducted monthly and is based on personal interviews with a national probability sample of 2500 homes. A female adult is interviewed in half of the homes, and a male adult is interviewed in the other half. Each AIM interview consists of two parts:

---

[11] "Roper Custom Question Service: A Unique Omnibus Research Facility," a brochure prepared (in 1977) by the Roper Organization, Inc. One Park Avenue, New York, N.Y. 10016.

[12] *AIM: A Monthly Omnibus Research Service Designed to Meet the Needs of Participating Clients,* a booklet published by R. H. Bruskin Associates, 303 George Street, New Brunswick, N.J. 08903.

1. The respondent is asked to examine a standard-size deck of 52 cards, each card containing a product name, slogan, trademark, symbol, or the like, submitted by a client firm. The purpose of this part is to measure consumer awareness and recognition of whatever is printed on each card.
2. The respondent is asked a set of custom questions—again, submitted by individual client firms—dealing with subjects such as product use, interest in products and services, and advertising or media exposure.

Research in Use 5.1 describes the two parts in greater detail, and Research in Use 5.2 contains examples of cards used in an AIM survey.

Responses to the card messages and/or questions submitted by each client are separately tabulated (by sex, age, income, and region) and made available only to that client. As in the case of the Roper Custom Question Service, the cost of AIM varies from client to client. It depends on the number of cards a client wants to buy (from the 52 cards available per survey), the type of material to be printed on each card, and the number and complexity of custom questions submitted.

Research in Use 5.1 *R. H. Bruskin Associates*

AIM: A Customized Syndicated Service

THE CARD SECTION

The AIM technique uses actual playing cards with the usual numbers and pictures replaced by the subscriber's test material. A full deck of 52 cards is used.

At the beginning of the interview, the respondent is handed the shuffled deck of cards containing all of the items to be tested. He then sorts these cards into two groups—those he knows or thinks he has seen or heard before, and those he is sure he has never seen or heard before (association). Additional questioning follows to determine the respondent's knowledge and identification of the items (identification).

Extensive testing and experience have revealed that this technique stimulates a great deal of respondent interest. The respondent feels he is "playing a game," which results in a high level of alertness and cooperation during the course of the entire interview. This psychologically favorable

(Continued)

Research in Use 5.1 *Continued*

climate allows for the highest level of identification and minimizes losses which might otherwise occur.

THE OMNIBUS SECTION

Upon completion of this first phase, the interviewer then proceeds with the omnibus section covering the special questions inserted by participating clients. Some clients have purchased as many as three pages of questions, others as little as one question.

In this section the client can insert questions of any type and on any subject. Many techniques have been employed—in-home audits; numerical rating scales; word and statement image listings; self-administered questions; card sorts; as well as the standard question procedures.

Visual aids can be used where appropriate to the objective of the question. The client's questions can follow in sequence or can appear in several positions in this omnibus section if desired.

The onmibus portion of the questionnaire is flexible. It is tailored to the individual needs of *each* participating client.

Source: Reproduced from *AIM: A Monthly Omnibus Research Service Designed to Meet the Needs of Participating Clients,* a booklet published by R. H. Bruskin Associates, 303 George Street, New Brunswick, N.J., pp. 3–4. Used by permission.

Strictly speaking, the data provided by a service such as AIM is primary rather than secondary. Nevertheless, they can still be labeled "syndicated data" because they are collected on a regular basis and the data collection costs are shared by many clients. Moreover, such a service is still not as flexible or versatile as a primary-research project conducted by a client firm on its own. For instance, even in the case of AIM, since it is designed to cater to multiple clients, there are practical limits on the number of cards and/or questions that can be sold to any one client.

Commercial marketing research firms are not the only sources of syndicated data. Supermarkets equipped with checkout scanners are also a potential source. Major supermarket chains, in addition to using scanner data as internal secondary data, are making them available as syndicated data to packaged-goods producers such as Procter and Gamble and Kraft.

Trade Associations. There are literally thousands of trade associations, each representing a group of organizations that share a common trade or line of business. Such associations include the National Association of College Stores, Retail Bakers of America, Canadian Roofing Contractors Association, Safety Helmet Council of America, Newpaper Ad-

Research in Use 5.2 *R. H. Bruskin Associates*

Examples from an AIM Card Deck

PUEBLO COLORADO 81009

A

CHANGING THE WAY
AMERICA COOKS

RP

SHEER ENERGY

RQ

TENDER CHUNKS

RG

SIX FLAGS

B

YOU LOVED IT AS A KID
YOU TRUST IT AS A MOTHER

RT

Source: R. H. Bruskin Associates, 303 George Street, New Brunswick, N.J. Used by permission.

vertising Sales Association, Luggage and Leather Goods Salesmen's Association of America, Inc., Roadside Business Association, and American Cultured Dairy Products Institute.[13] These examples, in addition to illustrating the diversity of associations, show that even narrowly defined fields have their own trade groups.

---

[13] Several directories listing various associations are available; see, for example, Nancy Yakes and Denise Akey, *Encyclopedia of Associations*, vol. 1 (alphabetical index; 1979), vol. 2 (geographic and executive index; 1980) (Detroit: Gale Research).

Most trade associations collect background data about their members and markets. They may also collect other types of data of potential interest to their members. Hence they can be valuable sources of secondary data, although a researcher planning to tap such sources may face certain constraints. The data available through associations will usually be quite general, and certain types of data, such as competitively sensitive data about an association's members or trade, may not be available at all. Another potential problem is that some trade associations will release data only to their members; thus whatever data exist will obviously be unavailable to an outsider.

Miscellaneous Sources. Miscellaneous sources include sources that do not neatly fit into the three previous categories. Examples are journals, magazines, research monographs, textbooks, and similar published materials. Most libraries, especially those affiliated with academic institutions, have numerous collections of such sources.

A distinguishing feature of sources in this category is that the data they contain is in the form of insights, ideas, suggestions, and so forth, rather than in the form of numbers. Thus their primary purpose is to stimulate the thinking of researchers and decision makers, as opposed to providing them with numbers they can analyze. For example, the *Journal of Marketing* article mentioned at the beginning of this chapter—the article suggesting a relationship between the tempo of background music and supermarket sales—offers supermarket managers food for thought. Similarly, articles dealing with sales force motivation may provide valuable clues to a sales manager who is wondering why his or her sales force's performance is significantly below the performance of competitors' sales forces selling similar products. A miscellaneous source that is particularly useful in understanding demographic trends is *American Demographics*, a periodical carried by many university libraries.

yht.vedot luettelot hakemistot

Abstracts, Directories, and Indexes. Abstracts, directories, and indexes, in contrast to the sources discussed earlier, are *guides* or *references* for identifying suitable external secondary-data sources. They are listings of available data sources, classified according to subject matter, topic area, and the like. Several examples of abstracts/directories/indexes are briefly described in the appendix to this chapter. Some of the guides in this category offer more than a mere listing of secondary-data sources; that is, they are① annotated and include brief outlines of or data summaries from other sources. Nevertheless, annotated guides are still secondhand sources. A researcher would do well to locate and examine the original source, even when a summary provided by an annotated guide appears to be adequate for the purpose. In short, abstracts, directories, and indexes are good starting points in the search for secondary data,

① varust. muist.panoilla

especially when a researcher has little idea about what secondary-data sources exist or what they have to offer.[14]

The rapid technological advances in the computer and communications fields are also revolutionizing the process of searching secondary-data guides to identify suitable data sources. Today many libraries and commercial firms offer ***computerized searches*** for locating potential sources of secondary data. Traditional, printed abstracts/directories/indexes are being replaced by electronic files. To conduct a computerized search, a researcher merely specifies certain key words relating to the topic area on which data are desired. A computer then searches the electronic files and culls all the sources containing the key words.

Computerized Search
A computer-aided approach for quickly identifying potential sources of secondary data by simply specifying certain key words relating to the topic area on which data are desired.

An obvious benefit of computerized searches is speed: What would take days or weeks for a traditional, manual search can be completed within minutes or hours by a computerized search. Moreover, the electronic files can be updated almost as soon as a new source of data becomes available, a feature printed guides do not have. Hence computerized searches can uncover the latest available sources.

Commercial firms specializing in computerized searches are reducing the time required for searching yet another way, namely, by providing direct communication links between their computers and their clients' computers. An example is FIND/SVP, a New York–based information and research firm. It recently introduced a service called FINDSEND, which, according to the firm's manager of information resources, "enables us to electronically deliver search results directly to a marketer's terminal, eliminating the need to wait for receipt of printouts by mail."[15]

The availability of fast, efficient, and thorough computerized searches makes it all the more critical that a researcher conduct a search for secondary-data sources before embarking on any form of external data collection. Conducting such a search not only saves time but also eliminates needless expenditures, as one of FIND/SVP's clients apparently discovered:

> A marketing researcher for a high-technology company was about to commission a $20,000 custom research study of the robotics market, a study which would take many months to complete.
>
> At the last minute, he called FIND/SVP to see if a similar study was commercially available. He was told that not one but several such studies were for sale and that virtually all the information wanted could be had immediately for a few thousand dollars.[16]

---

[14] For an excellent discussion of how to conduct an effective and efficient search for secondary data, see David W. Stewart, *Secondary Research: Information Sources and Methods* (Beverly Hills, Calif.: Sage, 1984).

[15] FINDSEND Service Transmits Results of Data-Base Searches to Marketer's Terminal," *Marketing News,* November 26, 1982, Section 1, p. 14.

[16] Bernie Whalen, "Off-the-Shelf Market Research Studies Save Time and Money," *Marketing News,* September 17, 1982, Section 2, p. 24.

MANAGING SECONDARY DATA

The proliferation of secondary data (including internal secondary data) presents both an opportunity and a problem to researchers and decision makers. The variety and richness of the available secondary data can often minimize or even eliminate the need for primary data. But just keeping abreast of all the available data, without being overwhelmed, is a challenge. Indeed, two expressions frequently heard in business circles these days are "information explosion" and "drowning in data."

Secondary-Data Management *Creating and operating a system for continuously monitoring various secondary-data sources and quickly retrieving needed data.*

Capitalizing on the so-called information explosion without drowning in data requires effective ***secondary-data management***—that is, creating and operating a system for continuously monitoring various data sources and quickly retrieving needed data. Effective secondary-data management is a critical component of a sound marketing information system. Many firms have established their own corporate libraries as a means of managing secondary data effectively. A well-equipped corporate library, with computerized data storage and retrieval capabilities, can be a valuable asset.[17]

Information Consultant *An individual with special skills pertaining to effective secondary-data management.*

Recognition of the need to manage secondary data effectively has given birth to a new breed of specialists in the marketing research field, namely, ***information consultants.*** Gary Bratton, vice president of Infosource, Inc., describes the role and the capabilities of information consultants as follows:

> Information research and analysis are an information consultant's stock in trade. The information consultant is constantly seeking new sources of information and evaluating their content for decision needs. . . . Most marketing executives have their favorite eight or ten sources of information with which to gather information. But there are literally hundreds that provide a much broader base from which to extract more exact info. . . . An information consultant's strength is knowing the contents of electronic files that are not specifically labeled "marketing." In most instances, sales and marketing people would not know how to identify these files, much less have the expertise to research them.[18]

The potential benefits of good data management are not limited to external secondary data. Systematically monitoring internal secondary data and using them in creative ways can also be very productive. A firm that effectively manages internal secondary data can quickly spot market opportunities and problems. Sears is an example. A well-maintained and -managed customer data base is apparently a major contributor to the firm's rapidly expanding service business. According to *Business Week,*

---

[17] James R. Fries, "Library Support for Industrial Marketing Research," *Industrial Marketing Management,* 11 (February 1982), pp. 47–51; see also "Putting the Library on a Computer," *Business Week,* March 30, 1981, pp. 104–106.

[18] "Information Specialists Can Tap Computer Data Bases to Provide Market Intelligence," *Marketing News,* November 27, 1981, Section 2, p. 18.

> By keeping track of the appliance purchases made by each customer, as well as the service plans and maintenance calls for these appliances, Sears has been able to create a powerful marketing tool that is helping it boost service revenue and win customer goodwill. This program is just one of the many ways the retailing giant is mining the wealth of information that it has accumulated on its millions of customers—data that will enable it to develop new products and services.[19]

SUMMARY

Secondary data, although collected for other purposes, may be helpful in a project. In general, secondary data can be obtained more quickly and with less cost than primary data. Secondary data also have a less obvious advantage in that they can sometimes offer insights in situations where a single organization cannot collect primary data.

A potential limitation of secondary data relates to relevance. Available secondary data may lack relevance for a given project for any of the following reasons: the units in which the data are reported may be different from the desired units of measurement; the categories in which the data are grouped may not match the requirements of the project; the data may not be current. However, sometimes a researcher may be able to convert the measurement units, redefine the data categories and rearrange the data, and/or update the data. A researcher should explore these data modification options before dismissing available secondary data as irrelevant.

Another potential drawback of secondary data relates to accuracy, especially the difficulty in verifying the trustworthiness of the data. Answers to the following interrelated questions will shed light on this issue: Who collected the data? Why were the data collected? How were the data collected? The original source of secondary data is more likely to provide the documentation necessary to answer these questions than is a secondhand source. Therefore, if at all possible, the original source should be consulted.

Secondary-data sources can be broadly divided into internal sources (i.e., those within an organization) and exernal sources (i.e., those outside an organization). The search for secondary data should begin with internal sources since obtaining internal secondary data will require far less time, effort, and money than external secondary data. They may also be more pertinent to the situation being researched.

There is a very wide variety of secondary-data sources. One useful way of classifying them is as follows: government sources; syndicated sources; trade associations; miscellaneous sources; and abstracts, directories, and indexes. Each of these categories has distinct features.

---

[19] "Business Is Turning Data into a Potent Strategic Weapon," *Business Week*, August 22, 1983, p. 92.

Government agencies collect a huge volume of data, much of which are free and readily accessible. Data offered by syndicated sources are in general more expensive than those offered by government sources, however, they may also be more relevant to a research project. Trade associations are good sources of background data about particular trades or professions, although such data may be too general for use by a researcher or unavailable to nonmembers. Miscellaneous sources—journal articles, research monographs, and textbooks—are capable of providing insights, hunches, or ideas pertaining to a research project. Abstracts, directories, and indexes serve as guides in identifying other sources of secondary data; recent technological advances have resulted in computerization of many abstracts, directories, and indexes.

Coping with the variety and abundance of secondary data available today is a major challenge. Effective management of secondary data is essential if maximum benefits are to be derived from them. Systematic procedures for identifying and scanning secondary-data sources and rapidly retrieving required data when needed are the key elements of effective secondary-data management. More and more firms are using information consultants, who specialize in productively utilizing various sources of secondary data.

KEY TERMS

Secondary data
Primary data
Syndicated data (services)
Original source
Secondhand source
Internal source
External source
Computerized search
Secondary-data management
Information consultant

QUESTIONS

1. Primary data should be collected only when suitable secondary data are not available." Discuss this statement.
2. What are the three factors that may reduce the relevance of secondary data? Give one example to illustrate each factor.
3. Suppose you have found secondary data that you feel are relevant to a particular problem you are facing. Describe further checks you would make before deciding to use the secondary data.
4. What is the difference between original and secondhand sources of secondary data?
5. "Internal secondary data should only be used when appropriate external secondary data are not available." Do you agree or disagree with this statement? Give reasons for your answer.
6. Discuss the advantages and disadvantages of secondary data obtained from government sources and those obtained from syndicated sources.

7. The inventor of a special type of package material for bread and other baked goods sold through retail outlets wants to estimate the market potential for the material. What specific types of secondary sources should this inventor look into? Why?

8. What does *managing* secondary data mean? Why is it important?

9. Using the appropriate secondary-data source(s), obtain the following information for the standard metropolitan statistical area (SMSA) in which you live or for the SMSA closest to you. For each piece of information, state the source(s) you used and describe any calculations you made.

a. Number of retail outlets and the annual sales per outlet.
b. Number of law firms.
c. Percentage of families with children under 18 years of age.
d. Identification of the census tracts with the highest and lowest per capita incomes and their geographic locations.
e. Proportion of the total number of residential housing units that are less than 10 years old.
f. Proportion of residential housing units with seven or more rooms in (i) the entire SMSA; (ii) the census tract with the highest per capita income; and (iii) the census tract with the lowest per capita income.

External Secondary-Data Sources

As was pointed out in the chapter, listing all potentially valuable external secondary-data sources is impossible. Therefore, this appendix will briefly describe selected sources to illustrate the following categories: government sources; syndicated sources; and abstracts, directories, and indexes. The purpose here is to give you a sense of the nature and the scope of the available sources. The included sources should be of interest to marketing researchers dealing with a variety of situations. However, going through this source list is not a satisfactory substitute for a comprehensive search conducted in a library, preferably with the help of an experienced reference librarian.

GOVERNMENT SOURCES

U.S. Bureau of the Census

Census of Agriculture. This census contains data on farms and farming gathered through censuses taken every five years. Data are reported by state and county and include number and size of farms, farm production and income, farm employment, and characteristics of farm operations.

Census of Government. Conducted every five years, this census contains statistics about state and local governments. Reported data include government operating costs and revenues, number of employees, and size of payroll.

Census of Housing. Conducted every 10 years, this census contains a wealth of data about the number and features of households in the United States. Reported data include type of structure, occupancy, number of rooms, and number of appliances. Detailed geographic breakdowns (e.g., by census tracts or city blocks) are provided for standard metropolitan areas. It is updated annually by means of the *Current Housing Report.*

Census of Manufacturers. Conducted every five years, this census provides data for each state under 450 different classes of manufacturing industries. The data include number of establishments in each class, assets, rents, employment, payroll, cost of materials, value of shipments, and capital expenditures. Data are updated by the *Annual Survey of Manufacturers.* Data on some manufactured products are updated more frequently through *Current Industrial Report.* Industry statistics are also arranged by SIC number for convenient referencing.[1]

Census for Mineral Industries. Conducted every five years, this census reports data on about 50 4-digit SIC numeral industries for variables similar to those covered under the Census of Manufacturers.

Census of Population. Conducted every 10 years, this census reports population characteristics for states, counties, cities, and towns. More detailed geographic breakdowns are available for standard metropolitan areas. Reported data include age, sex, marital status, race, education level, employment status, family size, income, and occupation. It is updated annually by the *Current Population Report.*

Census of Retail Trade. Conducted every five years, this census reports data by state, standard metropolitan statistical area (SMSA), and areas outside SMSAs for over a hundred types of retail establishments. It includes data on number of outlets, total sales, employment, and payroll. The *Monthly Retail Trade* updates some of these data.

Census of Service Industries. Conducted every five years, this census provides data, similar to those reported by the Census of Retail Trade, for service establishments such as automotive service centers, hotels, and law firms.

Census of Transportation. Conducted every five years, this census provides a variety of travel-related data. It includes three principal types of data: (1) truck inventory and use (e.g., number of vehicles, body type, and major application); (2) nonlocal passenger travel (e.g., number and characteristics of trips); and (3) commodity shipments (e.g., ton-miles and means of transport).

Census of Wholesale Trade. Conducted every five years, this census contains data on variables such as number of establishments, sales, operating expenses, types of operation, employment, and payroll. Data are reported separately for over one hundred fifty types of wholesaler categories and also are classified by counties and SMSAs. Some of the data are updated by the *Monthly Wholesale Trade.*

County Business Patterns. This census is an annual summary of data on number and types (SIC) of business establishments as well as their employment and taxable payroll. The data are broken down by industry and by county.

---

[1] The standard industrial classification (SIC) system was developed by the federal government to classify products, services, and industries. It uses unique code numbers called SIC codes. Many types of data collected by the government and other sources are classified by SIC codes. Details about the SIC system are available in *Standard Industrial Classification Manual* (Washington, D.C.: U.S. Government Printing Office, 1972).

Other Government Sources

Business Statistics. This report is published every other year by the U.S. Department of Commerce. It contains a historical series of data (from the year 1947 to date) on numerous variables monitored monthly by the *Survey of Current Business* (described later) and serves as a supplement to the *Survey of Current Business.*

Economic Indicators. This report is published once a month by the Council of Economic Advisors. It contains the most recent data on economic variables like personal consumption expenses, gross national product, wages, and prices. It is a good source for assessing general economic conditions.

Economic Report of the President. This report is the published version of the speech delivered by the president of the United States to the U.S. Congress regarding the state of the economy. It contains summary tables of economic data gathered by various other agencies.

Statistics of Income. This annual publication of the U.S. Internal Revenue Service consists of a series of three reports: (1) *Corporation Income Tax Returns,* (2) *Business Income Tax Returns,* and (3) *Individual Income Tax Returns.* The three reports contain data derived from tax returns filed by corporations, proprietorships/partnerships, and individuals. The data are broken down by states and SMSAs.

Survey of Current Business. This monthly publication of the Bureau of Economic Affairs of the U.S. Department of Commerce contains data on hundreds of variables related to gross national product, national income, personal income, personal expenditures, government receipts and expenditures, savings and investments, income and employment by industry, and foreign transactions. An annual issue of the *Survey,* called *National Income Issue,* is a valuable source of national income and account statistics.

U.S. Industrial Outlook. This annual publication of the U.S. Department of Commerce presents brief descriptions of and statistics related to recent trends and future projections for more than two hundred different industries. The areas covered include supply and demand, domestic and international developments, employment, and prices.

Government Sources of Foreign Statistics

The U.S. Department of Commerce, the United Nations, and the governments of foreign countries are good sources of international data related to people, markets, and business. Examples of publications from these three types of sources are outlined below.

U.S. Department of Commerce Publications. *Foreign Economic Trends and Their Implications for the United States* is a report published semiannually or annually; it is prepared by United States embassies in over a hundred foreign countries. *Overseas Business Reports* is actually a series of reports on individual countries published with varying frequencies by the Domestic and International Business Administration (DIBA) of the Department of Commerce. Many of those reports include data about industry trends, trade regulations, taxes, and the like, and list other sources of economic and commercial data in each country. *International Economic Indicators and Competitive Trends* is published every three months. It provides data on economic trends and indicators for seven leading competitors of the United States: Canada, France, Italy, Japan, the Netherlands, the United Kingdom, and West Germany.

United Nations Publications. *World Economic Survey* is published annually and reports data on production and trade in various countries. *Yearbook of International Trade Statistics* reports data on imports and exports of over one hundred countries during the past four years. *World Trade Annual* provides data on international trade, classified according to type of commodity as well as by country. *Statistical Yearbook for Asia and the Pacific* contains statistics related to population characteristics, agriculture, industry, transportation, prices, household expenditures, and other business and social factors. It is organized by country. The United Nations also publishes similar statistical reports for certain other parts of the world, including a compound volume called *Statistical Yearbook* covering all United Nations countries.

Other Publications. The Organization of Economic Cooperation and Development publishes a variety of reports, with varying frequency, on 24 member countries. Examples include *Main Economic Indicators, Foreign Trade Statistics Bulletins,* and *Industrial Production: Historical Statistics;* they are published in Paris. The Statistical Office of the European Communities (EUROSTAT) releases a series of reports (e.g., *General Statistics, Foreign Trade, Monthly Statistics, Industrial Statistics, Yearbook,* and *Social Statistics Annual*) covering nine countries belonging to the European Community. The statistics reported cover a broad range of economic, population, and social variables; the reports are published in Brussels. The government statistical offices of many individual countries (e.g., Canada, Great Britain, and India) collect and publish a variety of statistics on a quarterly, monthly, or weekly basis.

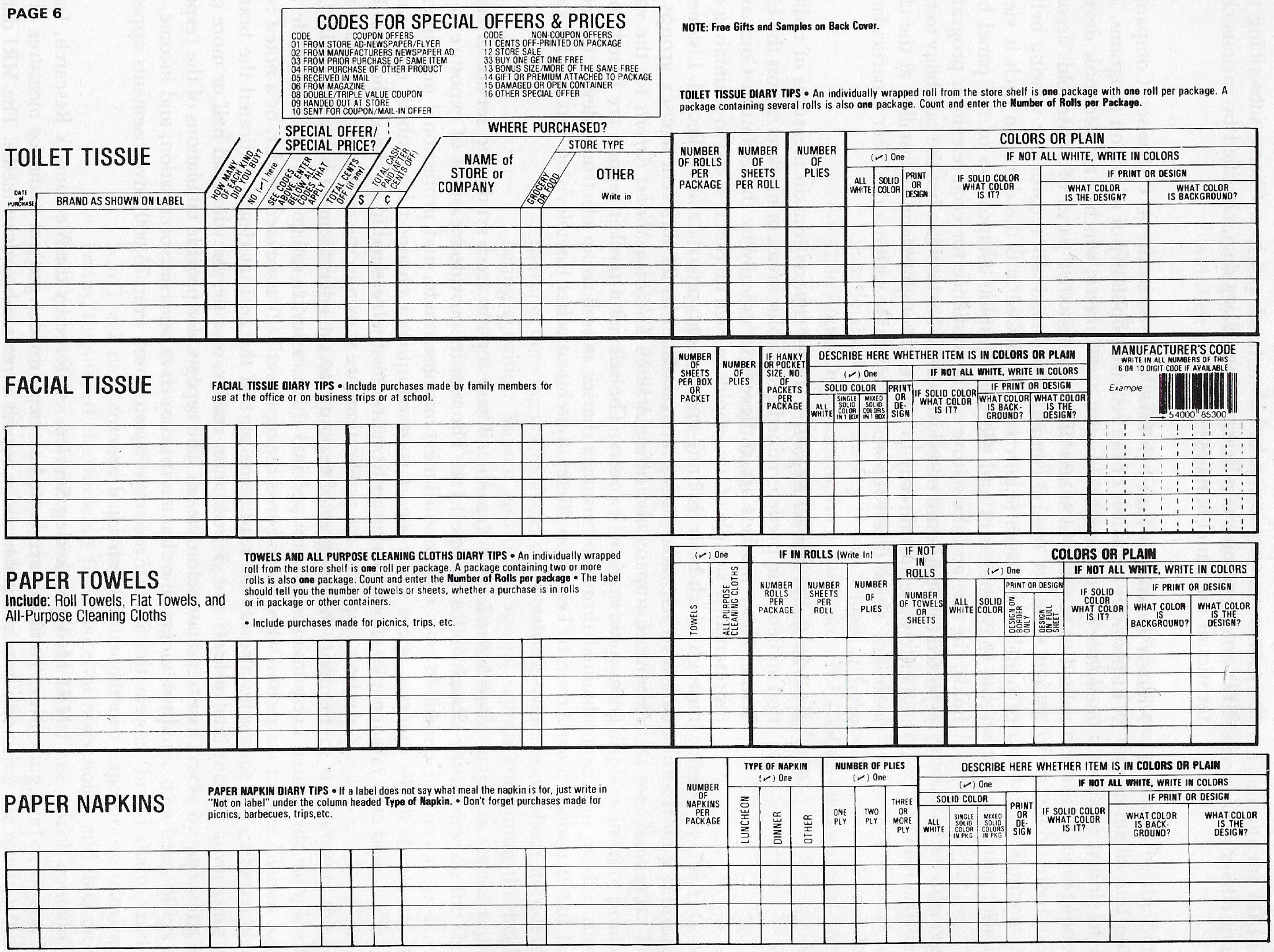
SYNDICATED SOURCES

One useful way of classifying syndicated sources is to group them according to the origin of data reported by them. Marketing research firms typically gather data for their syndicated services from people, retail outlets, wholesalers, or from a combination of units (including various other organizations and publications). Let us label these types of data "consumer data," "retail data," "wholesale data," and "miscellaneous data," respectively. Selected examples of syndicated sources under each of these categories are briefly described below.

Sources of Consumer Data

National Purchase Diary Panel (NPD). This service is offered by the NPD Research, Inc., of New York. Data are gathered from consumers on some fifty different product categories. The panel contains 13,000 families, representing a broad cross section of the United States. Panel members are asked to keep detailed records about the purchase and use of brands within each of the product categories. A special diary is provided by NPD for this purpose, and panel members mail completed diaries to the firm once a month. Exhibit A.5.1 contains

Exhibit A.5.1 *Sample Page from an NPD Diary*



PAGE 6

CODES FOR SPECIAL OFFERS & PRICES

CODE	COUPON OFFERS	CODE	NON-COUPON OFFERS
01	FROM STORE AD-NEWSPAPER/FLYER	11	CENTS OFF-PRINTED ON PACKAGE
02	FROM MANUFACTURERS NEWSPAPER AD	12	STORE SALE
03	FROM PRIOR PURCHASE OF SAME ITEM	33	BUY ONE GET ONE FREE
04	FROM PURCHASE OF OTHER PRODUCT	13	BONUS SIZE/MORE OF THE SAME FREE
05	RECEIVED IN MAIL	14	GIFT OR PREMIUM ATTACHED TO PACKAGE
06	FROM MAGAZINE	15	DAMAGED OR OPEN CONTAINER
08	DOUBLE/TRIPLE VALUE COUPON	16	OTHER SPECIAL OFFER
09	HANDED OUT AT STORE		
10	SENT FOR COUPON/MAIL-IN OFFER		

NOTE: Free Gifts and Samples on Back Cover.

TOILET TISSUE DIARY TIPS • An individually wrapped roll from the store shelf is **one** package with **one** roll per package. A package containing several rolls is also **one** package. Count and enter the **Number of Rolls per Package.**

TOILET TISSUE

DATE of PURCHASE | BRAND AS SHOWN ON LABEL | HOW MANY OF EACH KIND DID YOU BUY? | SPECIAL OFFER/ SPECIAL PRICE?: NO (✓) here; SEE CODES ABOVE, ENTER BELOW ALL CODES THAT APPLY; TOTAL CENTS OFF (if any); TOTAL CASH PAID (AFTER CENTS OFF) $ ¢ | WHERE PURCHASED?: NAME of STORE or COMPANY; STORE TYPE: GROCERY OR FOOD; OTHER Write in | NUMBER OF ROLLS PER PACKAGE | NUMBER OF SHEETS PER ROLL | NUMBER OF PLIES | COLORS OR PLAIN: (✓) One: ALL WHITE, SOLID COLOR, PRINT OR DESIGN; IF NOT ALL WHITE, WRITE IN COLORS: IF SOLID COLOR WHAT COLOR IS IT?; IF PRINT OR DESIGN: WHAT COLOR IS THE DESIGN?, WHAT COLOR IS BACKGROUND?

FACIAL TISSUE

FACIAL TISSUE DIARY TIPS • Include purchases made by family members for use at the office or on business trips or at school.

NUMBER OF SHEETS PER BOX OR PACKET | NUMBER OF PLIES | IF HANKY OR POCKET SIZE, NO. OF PACKETS PER PACKAGE | DESCRIBE HERE WHETHER ITEM IS **IN COLORS OR PLAIN**: (✓) One: SOLID COLOR: ALL WHITE, SINGLE SOLID COLOR IN 1 BOX, MIXED SOLID COLORS IN 1 BOX; PRINT OR DE-SIGN; IF NOT ALL WHITE, WRITE IN COLORS: IF SOLID COLOR WHAT COLOR IS IT?; IF PRINT OR DESIGN: WHAT COLOR IS BACK-GROUND?, WHAT COLOR IS THE DESIGN? | MANUFACTURER'S CODE WRITE IN ALL NUMBERS OF THIS 6 OR 10 DIGIT CODE IF AVAILABLE

Example 54000 85300

PAPER TOWELS

Include: Roll Towels, Flat Towels, and All-Purpose Cleaning Cloths

TOWELS AND ALL PURPOSE CLEANING CLOTHS DIARY TIPS • An individually wrapped roll from the store shelf is **one** roll per package. A package containing two or more rolls is also **one** package. Count and enter the **Number of Rolls per package** • The label should tell you the number of towels or sheets, whether a purchase is in rolls or in package or other containers.

• Include purchases made for picnics, trips, etc.

(✓) One: TOWELS, ALL-PURPOSE CLEANING CLOTHS | IF IN ROLLS (Write In): NUMBER ROLLS PER PACKAGE, NUMBER SHEETS PER ROLL, NUMBER OF PLIES | IF NOT IN ROLLS: NUMBER OF TOWELS OR SHEETS | COLORS OR PLAIN: (✓) One: ALL WHITE, SOLID COLOR, PRINT OR DESIGN: DESIGN ON BORDER ONLY, DESIGN ON FULL SHEET; IF NOT ALL WHITE, WRITE IN COLORS: IF SOLID COLOR WHAT COLOR IS IT?; IF PRINT OR DESIGN: WHAT COLOR IS BACKGROUND?, WHAT COLOR IS THE DESIGN?

PAPER NAPKINS

PAPER NAPKIN DIARY TIPS • If a label does not say what meal the napkin is for, just write in "Not on label" under the column headed **Type of Napkin.** • Don't forget purchases made for picnics, barbecues, trips, etc.

NUMBER OF NAPKINS PER PACKAGE | TYPE OF NAPKIN (✓) One: LUNCHEON, DINNER, OTHER | NUMBER OF PLIES (✓) One: ONE PLY, TWO PLY, THREE OR MORE PLY | DESCRIBE HERE WHETHER ITEM IS **IN COLORS OR PLAIN**: (✓) One: SOLID COLOR: ALL WHITE, SINGLE SOLID COLOR IN PKG, MIXED SOLID COLORS IN PKG; PRINT OR DE-SIGN; IF NOT ALL WHITE, WRITE IN COLORS: IF SOLID COLOR WHAT COLOR IS IT?; IF PRINT OR DESIGN: WHAT COLOR IS BACK-GROUND?, WHAT COLOR IS THE DESIGN?

Source: NPD Research, Inc., 900 West Shore Road, Port Washington, N.Y. 11050. Used by permission.

one page from the NPD diary. The richness of the NPD data is evident from the numerous questions asked about each product. Clients subscribing to the NPD service receive tabulations of the diary data, classified by various characteristics of the panel families.

Nielsen Television Index (NTI). Offered by the A. C. Nielsen Company, the NTI provides measures of the relative popularity of television programs. These measures are obtained through an audimeter, which is an electronic device attached to the television sets of a representative panel of households across the United States. A central computer automatically records the viewing behavior of panel households (e.g., channels viewed and the duration each is viewed). From these measures and the demographic composition of the panel households, the size and the nature of the audience for each television program are reported to clients once every two weeks. Nielsen also has the capability to report audience measures directly to clients through terminals installed at the clients' locations. These direct reports are available within a day after a particular program is aired.

Arbitron Radio Market Reports. These reports contain measures of radio station audiences in predetermined geographic areas, or markets, across the United States. These markets are surveyed periodically by the Arbitron Company. A representative sample of households in each market is asked to maintain a diary, through which basic listening and demographic data are collected. Listenership data for each radio station are tabulated by 21 different age/sex groups. The Arbitron Company also offers *Arbitron Television Market Reports* for the various geographic markets it covers. The television market reports are based on data obtained through consumer diaries as well as an electronic device (called the Arbitron Television Meter System) attached to television sets in the sample households.

Starch Readership Reports. These reports are a syndicated service offered by Starch INRA Hooper, Inc. The reports provide measures of exposure to print ads in over ninety consumer, business, trade, and professional magazines and newspapers. Each report is based on data obtained from about one hundred people, carefully chosen to represent the readership of the publications being studied. Copies of the publications are shown to the respondents, who are asked to report their degree of exposure to the advertisement being researched. The respondents are grouped into three exposure categories: (1) noted reader, or a person who remembered seeing the ad; (2) associated reader, or a noted reader who also saw or read some part of the ad that clearly indicated the brand or advertiser; and (3) read most reader, or a person who read half or more of the written material in the ad. The reports also provide tabulations of the responses by demographic characteristics such as age, sex, occupation, and income. Each year the *Starch Readership Reports* cover over 75,000 advertisements appearing in about 1000 individual issues.

MRI Mediamarketing Service. Offered by the Mediamark Research, Inc., of New York, this syndicated service provides general audience measures of over two hundred publications as well as certain broadcast media. The MRI data are collected through two waves of personal interviews a year. In each wave a representative sample of 10,000 people are interviewed about their media exposure habits and their use of various products and services. Therefore the MRI reports provide useful information about the personal characteristics of media audiences and their purchase behavior.

Simmons Media/Marketing Service. This service, provided by the Simmons Market Research Bureau, Inc., also provides data on the media and purchase habits of consumers. The data are collected from a national sample of 15,000 individuals and relate to their exposure to media (magazines, newspapers, television, and radio) as well as their purchase and use of several hundred products. The media, product, and demographic data provided by the Simmons Media/Marketing Service can be cross-tabulated to yield useful marketing insights.

Sources of General Consumer Attitudes. Several syndicated services regularly monitor and report the mood of the public with respect to general issues such as social, economic, and political issues. Well-known examples of these services include the Harris Survey (Louis Harris and Associates, Inc.), Roper Reports (The Roper Organization, Inc.), the Gallup Poll (Gallup Organization, Inc.), and the Yankelovich Monitor (Yankelovich Clancy Shulman).

Sources of Retail Data

Nielsen Retail Index. This widely used service is offered by the A. C. Nielsen Company. Data are regularly collected by field auditors from a representative sample of 2500 supermarkets, drug stores, and mass merchandisers. The collected data include inventory levels, prices, special displays, in-store promotions, purchases by the retailers, and sales to consumers for various product categories and brands sold through these stores. While the basic Nielsen Retail Index is useful in deriving brand shares and obtaining information about store-level marketing activities, Nielsen also offers special reports combining the field audit data with relevant data from other sources. For example, a special report can be prepared to study the effect of national media advertising on the shares of the advertised brands in different markets. Also, Nielsen offers an on-line, time-sharing computer system called INF*ACT, through which clients can directly access the field audit data stored in Nielsen's computer data base.

Nielsen National ScanTrack Service. This service supplements the Nielsen Retail Index and is based on data obtained from a nationwide sample of high-volume, scanner-equipped supermarkets. Unlike data for the Nielsen Retail Index, product and price data for the ScanTrack service are gathered electronically; however, data on such things as displays and retailer advertising are obtained through weekly in-store observations. ScanTrack, by providing weekly reports or even more frequent access to its computer data base, offers an opportunity for almost continuous monitoring of sales, brand shares, and so on. ScanTrack data are now supplemented by data on exposure to television programs, monitored electronically for a panel of households, and data on household characteristics. The resulting *single-source data base,* which contains in one place data on supermarket purchases, television viewing, and demographics for a representative sample of households, is revolutionizing the way in which packaged-goods marketers access and use information about their markets.

BehaviorScan. This computer-based service is offered by Information Resources, Inc. (Chicago), and is similar to Nielsen's ScanTrack service. BehaviorScan also offers a single-source data base by electronically tracking the supermarket purchases and television exposure of a panel of households. Each panel household is issued a coded plastic card, which is scanned along with the household's purchases so that a central computer is able to keep track of who is buying what. (BehaviorScan is described in greater detail in Chapter 2.)

ScanAmerica. A service similar to ScanTrack and BehaviorScan offered by SAMI-Burke.

Audits and Surveys National Total Market Index. This service is provided by Audits & Surveys, Inc., of New York. The data for this service are obtained through field audits of retail outlets, as in the case of the Nielsen Retail Index Service; however, a distinguishing feature of this service is that it is a product audit rather than a store audit. Specifically, retail sales, inventory levels, prices, and so forth, for a particular product are monitored in all types of stores selling the product category. This service can be especially helpful in gaining insights about geographic differences in sales, prices, types of outlets, and distribution coverage for individual products and brands.

Sources of Wholesale Data

Selling Areas—Marketing, Inc. (SAMI). This source is perhaps the best known and most widely used source of syndicated data gathered at the wholesale level. However, SAMI data are obtained only from certain types of wholesalers; they report the movement of dry grocery products, general household supplies, frozen foods, health and beauty aids, and some refrigerated products. Excluded from SAMI reports are perishable products like milk, bread, and fresh meat. However SAMI reports do cover the warehouse movement of products accounting for over 80% of all food sales. The share of each brand is computed from brand volumes withdrawn from the warehouses and is reported once every four weeks.

Sources of Miscellaneous Data

We defined miscellaneous syndicated data as data put together by private organizations from a variety of sources, including published sources. However, this category is not necessarily unique, because some of the examples provided here can also be classified under the next category of sources, namely, abstracts, directories, and indexes.

PRIZM (Potential Rating Index by Zip Market). This data base is prepared by the Claritas Corporation. It is the result of combining demographic data from the Census of Population and Housing and life-style/attitude data obtained through various regional surveys. The PRIZM contains descriptive profiles of residents in each of 36,000 five-digit zip code areas in the United States. The

PRIZM data base has been combined with various other syndicated services (e.g., the MRI Mediamarketing Service discussed earlier) to produce other special data bases.[2]

Predicasts. Released quarterly by Predicasts, Inc. (Cleveland, Ohio), this report is a very useful source of short-term and long-term forecasts for various industries and for individual products. The product forecasts are classified by SIC number. Complete references of the variety of data sources used in arriving at the forecasts are included in each report. Predicasts, Inc., also publishes *Worldcasts,* which report similar quarterly forecasts for a number of foreign countries.

Survey of Buying Power. This annual volume is released by the publishers of *Sales & Marketing Management* magazine. Since it is available in most university libraries, it can be readily accessed. The *Survey* contains data on variables such as population, retail sales, and income for various states, counties, cities, and SMSAs; it also reports an index of buying power for each market for which data are provided.

ABSTRACTS, DIRECTORIES, AND INDEXES

Listings under this category are perhaps more important than the illustrative sources outlined above, because abstracts, directories, and indexes are sources of sources: They are the logical starting point for conducting a comprehensive secondary-data search. The following list, although not exhaustive, contains some of the more commonly used government and private guides.

Government Guides

Guide to U.S. Government Statistics. Compiled by John L. Andriot (McLean, Va.: Documents Index, 1973), this guide contains an annotated list of government statistical publications, arranged according to the publishing agency. Another useful source of government statistical publications is the *American Statistical Index,* issued monthly by the U.S. Congressional Information Service and published annually as a cumulative volume.

Index to International Business Publications. Published by the U.S. Domestic and International Business Administration (DIBA) of the U.S. Department of Commerce, this publication lists all DIBA studies classified under the following categories: export market reports, foreign economic trends, global market surveys, and overseas business reports.

Monthly Catalog of United States Government Publications. Sold by the Superintendent of Documents (U.S. Government Printing Office, Washington, D.C.

---

[2] Numerous commercial research firms offer special services related to and based on data about consumer demographics. A useful listing of such firms, along with the services they offer, is provided by Martha Farnsworth Riche, "Data Companies 1983: Directory of Data Firms," *American Demographics,* February 1983, pp. 28–39.

20402), this monthly publication classifies listings in four different ways: by authors, by title, by subject, and by series/report. The U.S. Government Printing Office also publishes cumulative volumes semiannually and annually.

Statistical Abstract of the United States. Published annually since 1878 by the U.S. Bureau of the Census, this publication is designed to serve as a convenient volume of statistical references and as a guide to other statistical sources. It contains summary statistics (derived from the most recent censuses) on the social, political, and economic organizations of the United States. This publication also has a *Guide to Sources of Statistics and Publications* and a *Guide to State Statistical Abstracts.*

Private Guides and Online Data Bases

ADSEARCH. This file is a computerized index of all advertisements of a quarter page or larger appearing in 150 major consumer magazines in the United States. ADSEARCH is offered by Lockhead's DIALOG Information Service. It can be accessed and searched directly from a terminal in a client's location. The access charge is around $100 per hour. The advertisements included in ADSEARCH are indexed by product name, company, and ad characteristics. The index is updated every month. Lockhead's DIALOG Information Service also provides access to several other computerized indexes including MANAGEMENT CONTENTS (an index of articles appearing in 200 scholarly and academic business periodicals) and PREDICASTS TERMINAL SYSTEM (containing various data bases and indexes compiled by Predicasts, Inc.).

Business Information Sources. Compiled by Lorna M. Daniells (Berkeley: University of California Press, 1976), this publication is a very thorough index of data sources. The volume offers annotated coverage of government and private sources of domestic and international data.

Business Periodicals Index. Published monthly (except in August) by H. W. Wilson Company of New York, this widely used index lists articles appearing in numerous English language periodicals. The articles are grouped according to specific subject areas, which themselves are cross-referenced. A cumulative index is also published annually.

Conference Board Cumulative Index. This index lists various research studies conducted over the past 20 years by the Conference Board (New York), an international network of leaders who exchange information on management, economic, and public policy issues. The research studies cover a wide variety of areas including business economics and marketing.

Dun & Bradstreet Million Dollar Directory. Published annually by Dun & Bradstreet, Inc., of New York, this directory consists of three volumes and is a comprehensive listing of business firms. Sales revenues, number of employees, four-digit SIC numbers indicating lines of business, and names and addresses of top executives are provided for each firm.

Directory of Online Databases. This directory is published by Cuadra Associates of Santa Monica, California. It is updated quarterly and describes about 300 online data bases.

Encyclopedia of Information Systems and Services. This volume (6th edition) is compiled by Anthony T. Kruzas and Linda Varekamp Sullivan (Detroit: Gale Research, 1985). It contains a detailed listing of several thousand organizations engaged in creating, updating, and/or making available a variety of data bases.

F&S Index. Released weekly by Predicasts, Inc., this index is a good source of recent articles about companies and industries. The *F&S Index* is cumulated monthly as well as yearly. Predicasts also issues a monthly *F&S Index International* listing articles about foreign companies and industries; quarterly and annual cumulations are also available. Both the *F&S Index* and the *F&S Index International* can be assessed electronically through the PREDICASTS TERMINAL SYSTEM.

Social Sciences Citation Index. Published every four months by the Institute for Scientific Information of Philadelphia, this index lists all articles in 1400 social science periodicals and selected articles from 1200 periodicals in other disciplines. A cumulative index is also published annually.

Where to Find Business Information: A Worldwide Guide for Everyone Who Needs the Answers to Business Questions. This guide (2nd ed.) is compiled by David M. Brownstone and Gorton Carruth (New York: Wiley, 1982). It is intended "to show precisely and quickly where current business information is located and how to get it." It briefly describes the kind of information each source contains and how much it costs. Over 500 English language publications from around the world are included. The volume is divided into three sections: "Source Finder" (in which business topics are classified under 2500 subject areas), "Sources of Business Information" (which contains the actual reference material), and "Publisher's Index" (which classifies publishers of the various items covered by the guide).

CHAPTER 6

Overview of Primary-Data Collection Methods

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *State and illustrate the advantages and disadvantages of the questioning approach over the observation approach.*
- *State and illustrate the four different formats for questionnaires.*
- *Define projective technique and provide marketing applications for the following projective techniques: word association test, sentence completion test, thematic apperception test, and cartoon test.*
- *Summarize the advantages and disadvantages of the three basic questionnaire administration methods.*
- *State the distinct features and primary benefits of computer-assisted telephone interviewing and shopping mall intercept interviewing.*
- *Describe, with an example, each of the five dimensions along which observation methods can vary.*
- *Provide several illustrations of mechanical observation devices and state the purpose for which each is used.*
- *Select the most appropriate primary-data collection method given the characteristics of and constraints in a situation wherein data are to be collected.*

Primary Data
Data collected specifically for a project.

Data collected specifically for a research project are called ***primary data.*** Primary-data collection is necessary in situations where secondary data are either unavailable or inappropriate because of lack of relevance and/or accuracy. The purposes of this chapter are to review the basic methods of obtaining primary data, to discuss their pros and cons, and to present guidelines for choosing among them.

VARIETY OF DATA COLLECTION METHODS

Primary data can be collected through a number of different methods. In some situations several data collection options may be available for tackling the same research problem. To illustrate, consider Kwality Knitware, a national chain of clothing stores, which wants to evaluate the effectiveness of a special point-of-purchase promotion for its Kwality brand men's shirts. The promotion consists of a decorative display in which Kwality brand shirts are prominently and attractively featured. The following methods illustrate the ways Kwality Knitware can generate data for evaluating the effectiveness of the special display.

Method A. Conduct a brief personal interview of a sample of store customers during the promotional period. Ask the customers whether they bought Kwality brand men's shirts, and, if so, what motivated them to make the purchase. Also, ask them specific questions about their reactions to the special display.

Method B. At the conclusion of the special promotion, conduct a telephone survey of residents in the store's trading area to ascertain whether they visited the store during the promotional period, and, if so, what their reactions and responses were to the special display.

Method C. Same as method B, except that instead of conducting a telephone survey, mail questionnaires to a sample of residents, along with stamped return envelopes for sending back completed questionnaires.

Method D. Hire someone to observe customers and record their reactions as they pass by the special display. Ask the observer to record such things as whether customers stop to look at the display, how long they spend at the display, how interested they appear to be in it, and so on.

Method E. Videotape the area where Kwality brand shirts are featured to generate a continuous record of customer reactions and behavior as they approach and pass by the special display.

Method F. Program the store's electronic cash registers to automatically keep track of the total number of Kwality brand men's shirts sold during the promotional period.

Questioning *A primary-data collection approach in which respondents play an active role owing to their interaction with the data gatherer.*

Observation *A primary-data collection approach in which respondents play little or no active role in terms of interacting with the data gatherer.*

Methods A, B, and C involve questioning customers, while methods D, E, and F involve observing customers or their purchases. ***Questioning*** and ***observation*** are the two broad approaches available for collecting primary data. Numerous variations of these approaches are available, some of which are illustrated by methods A through F. Several other forms of questioning and observation will be discussed subsequently. The key distinction between the two stems from the role potential respondents play in the data collection process. In the questioning approach respondents play an active role because of their interaction or communication with the researcher. Indeed, some textbooks label the questioning approach as the communication approach, to distinguish it from observation, in which respondents do not directly interact or communicate with the researcher.[1]

Questionnaire *An instrument used for eliciting and recording responses in many, but not all, research projects employing the questioning approach.*

saada esin

The use of the term *questioning approach* rather than *questionnaire approach* in the foregoing discussion is deliberate, because not all questioning or interviewing situations involve the use of formal questionnaires. Stated differently, a ***questionnaire*** is simply a tool that is used in many, but not all, research projects employing questioning. For instance, while a large-scale consumer survey invariably requires a formal questionnaire, an informal interview of a few knowledgeable persons does not, despite the fact that both fall under the questioning approach.

The various questioning and observation methods are not necessarily restricted to certain types of research. Chapter 4 described three major research types: exploratory, descriptive, and experimental, where the latter two are forms of conclusive research. Either of the two data collection methodologies can be used in any of these research types. To illustrate, consider methods A through F for the Kwality Knitware case. Methods similar to each of these can be employed in *any* of the following types of research:

---

[1] See, for example, Gilbert A. Churchill, Jr., *Marketing Research: Methodological Foundations*, 4th ed. (New York: Dryden Press, 1987), p. 224.

- Exploratory research. Management wants to gain some initial insights about the effectiveness of the special point-of-purchase display. Initial insights can be obtained by gathering data through questioning or observation (in a fairly informal and flexible fashion) from just one or a few stores.
- Descriptive research. Management wants information of a specific nature, such as, What types of customers are attracted to the special display? How do customers who purchase Kwality shirts during the special promotion differ from those who do not? This information can be generated by simply making the data collection process more formal and specific and by collecting data from a larger, more representative sample of stores.
- Experimental research. Management wants to know if the special display *causes* a significant impact on customer perceptions and purchases of Kwality shirts. This research question can be answered by collecting data under more controlled conditions than exist in the previous (descriptive-research) setting. For instance, data on customer reactions and purchases can be gathered from a sufficiently large, representative sample of stores having the special display (i.e., an experimental group) and of stores not having the special display (i.e., a control group).

Thus while the formality/flexibility of the data collection process, the nature of the sample, and the conditions under which data are collected vary from exploratory to descriptive to experimental research, the same basic data collection methods can be used under each. The next section examines the relative advantages and limitations of questioning and observation.

QUESTIONING VERSUS OBSERVATION

To put the discussion in this section into proper perspective, we must mention at the outset a couple of points concerning questioning and observation. First, each approach is not necessarily a *substitute* for the other, because each has certain unique capabilities. For instance, in ascertaining young children's reactions to such stimuli as toys and educational programs, observation may be the only meaningful way of gathering the required information. Research in Use 6.1 illustrates the power and benefits of observation in such circumstances.

Research in Use 6.1 *Children's Television Workshop*

Application of Observation Techniques in "Sesame Street" Programming

Children's Television Workshop (CTW), which produces the widely acclaimed "Sesame Street" program, used imaginative observation techniques in determining the most effective format for the program.

Robert Waterman, in his book *The Renewal Factor*, describes the process CTW used to evaluate "Sesame Street" programming:

> The main worry [in ascertaining young children's response to the programming] all along was getting and holding the attention of this eminently distractible segment of the market. How to listen to that market, let alone do market research? After a few false starts [the researchers] settled on a technique that puts children in day-care centers in front of both a television monitor and a screen for carousel slide projection. As various segments of proposed *Sesame Street* programming are shown, slides designed to distract attention are shown alongside. Any time the child's attention shifts from the program to the slides, CTW knows it has a problem. They found, for example, that segments showing adults talking usually score low. As a result, a very creative idea, "The Man From Alphabet," which showed a bumbling kind of hero struggling with the alphabet, was dropped. He talked too much. . . .
>
> Other ideas were tested and survive to this day: Big Bird, a child surrogate who is clumsy, misunderstands, makes mistakes, and is incredibly naïve; Bert, who plays straight man to Ernie, whose childlike behavior is always bugging him.
>
> Cooney [founder and chief executive of CTW], who had been watching children's reactions to "regular" television, noticed that the commercials were generally more exciting to the kids than the shows. Preschoolers seemed to perk up during commercial time. Thus, the heavy educational parts of Sesame Street are presented as short-burst commercials, rather than as part of the "normal" programming: "We'll be right back after this message about the letter 'A'."

Source: Robert H. Waterman, Jr., *The Renewal Factor* (New York: Bantam Books, 1987), p. 152.

Second, although each approach has certain advantages over the other (as will be discussed shortly), those advantages may not hold true in every situation calling for primary-data collection. In other words, neither approach is likely to be *always* better than the other along each of the following dimensions.

Nonverbalisms

Versatility

A major advantage of the questioning approach is its versatility in the types of data it is capable of generating. The observation approach, by definition, is limited to collecting data about visible characteristics or variables. For example, the amount of time a customer spends looking at the special display for Kwality shirts can be recorded by human observers (method D) or by videotaping equipment (method E). But one cannot observe very well whether the customer is impressed, disappointed, annoyed, or whatever, during the time he or she is looking at the display. In contrast, any of the questioning methods (A, B, or C) can be used to ascertain how the customer felt about the display. In short, observation can provide data only on overt characteristics and behavior. By contrast, direct questioning can shed light on overt variables and also provide data on respondents' feelings, motives, intentions, and other unobservable variables. However, as we saw earlier, for certain special groups of respondents (e.g., young children) the questioning approach may not be viable.

Time and Cost

The questioning approach usually also has an edge over observation in terms of the time and cost needed for data collection, because a researcher will have much more flexibility in the data collection process when data are collected through questioning. In the Kwality Knitware case the researcher will have some latitude in data collection in each of the three questioning methods—personal interviews (method A), telephone interviews (method B), or mail survey (method C). Therefore these methods can be scheduled in such a way as to generate the needed data efficiently. In contrast, employing a human observer (method D) or using videotaping (method E) is likely to be time-consuming and expensive. The observer and videotaping equipment will not be recording anything relevant when there are no customers in the special-display area. This unproductive observation, in addition to prolonging the duration of data collection, will also have to be paid for.

What about the relative data collection efficiency of method F, namely, programming the electronic cash registers to keep track of Kwality shirt sales during the promotional period? This method is not likely to be more time-consuming and expensive than the various questionnaire methods because the total sales of Kwality brand shirts can simply be read from the cash registers with relatively little effort or expense. An important implication emerges here: *Not all* observation methods may be more expensive and time-consuming than questioning methods in a given situation. Alternatively, the questioning approach is usually, but not always, more efficient in terms of data collection speed and cost. In fact, recent advances in electronic observation technology are giving marketers who effectively use such technology a distinct advantage over competitors who rely on other means of obtaining feedback from their markets. Research in Use 6.2 illustrates a case in point.

Method F, however, has one limitation that may nullify its data collection efficiency: The only type of data method F can generate is the

Research in Use 6.2 *The Limited*

Effective Use of Electronic Observation

The Limited is a chain of retail stores that sells a variety of fashion apparel. The fashion apparel industry is highly competitive and volatile. Success in this industry is dependent on consumer whims and fancies that can change rapidly. Therefore speedy identification of the most popular fashions in a given season and having those fashions in stock before consumer preferences change are extremely critical. However, the time it takes for many firms in this industry to identify a potentially popular fashion and have it in stock is six months or more. On a number of occasions, this time interval is just too long, because consumer preference for the ordered fashion merchandise weakens considerably by the time the merchandise reaches the stores. In contrast, The Limited is able to predict accurately the most popular fashions and have them in stock *within 60 days*. It does so by combining the power of electronic observation with modern communication and information technologies. How The Limited has been able to cut the order-cycle time from over 6 months to 60 days is summarized below.

Point-of-sale computers installed in The Limited's retail outlets transmit information daily to company headquarters in Columbus, Ohio, regarding what is selling well and what is not. At headquarters, this information is interpreted by management and the most popular fashions, which need to be restocked rapidly, are identified. Orders for these items, along with facsimile illustrations of the items, are sent by satellite to suppliers' plants in the United States, as well as in Hong Kong, South Korea, Singapore, and Sri Lanka.

Apparel manufactured abroad is sent to a common collection point in Hong Kong. The clothing begins to collect in Hong Kong within a matter of days after orders are received from Columbus. About four times a week a chartered Boeing 747 rushes the goods to a highly automated distribution center in Columbus. There, the apparel is sorted, priced, and prepared for shipment within 48 hours of receiving it. Finally, the apparel is shipped by truck and plane to The Limited's over 3000 stores.

Obviously, electronic observation of sales at the retail level is the starting point and hence foundation for The Limited's ability to have the *right* apparel (the most popular fashions), at the *right place* (The Limited's retail outlets), at the *right time* (within 60 days after the apparel's popularity is sensed by management).

Source: This illustration is based on information presented in Jeremy Main, "The Winning Organization," *Fortune*, September 26, 1988, pp. 50–60.

sales of Kwality brand shirts. Hence it will be inadequate if Kwality Knitware's management wants richer data—such as the data capable of being generated through observation methods D and E—about customers' reactions and behavior in response to the special display.

Data Accuracy

Data accuracy—that is, the extent to which the collected data are error-free and trustworthy—is crucial from a managerial standpoint since it can significantly influence the soundness of decisions based on marketing research. Inaccurate data can be misleading and do more harm than good. How does the questioning approach compare with observation in terms of data accuracy? The answer depends somewhat on the characteristics of a specific research project, including its purpose and data requirements.

To illustrate, suppose Kwality Knitware's management is interested in determining the impact of the special display on sales of Kwality brand shirts. Questioning customers about their reactions to the display and their purchases of Kwality brand shirts may not produce accurate data. For one thing, respondents may be unable to accurately recall their reactions and purchases, especially in the case of a mail survey (method C) owing to the time lag involved. For another, respondents may be unwilling to reveal their true reactions or purchases; if so, they may refuse to answer questions or, worse still, give erroneous answers. How easy it is to get inaccurate data through the questioning approach is illustrated by the following example.

> **Example.** In the early eighties, several research firms specializing in the computer industry made predictions about the growth of the personal computer market which turned out to be overly optimistic. For instance, Info Corp. predicted in 1982 that the personal computer market would have sales of $21 billion by 1986. Yet by mid-1985, sales were well below $10 billion. Another firm, Future Computing, predicted in 1983 that 23% of all homes would have personal computers by 1985 in contrast to an actual ownership rate of 13% by mid-1985. A reason believed to be at least partly responsible for these erroneous forecasts is inaccurate survey data collected by the research firms. According to one computer distributor who participated in such surveys: "Everybody lies. The market researchers called the vendors and asked what they sold. When they called me, I tripled my sales because I knew my competitors were doing the same. In this business, it is a source of shame to admit your true numbers. The problem is, everybody starts to believe his own hype."[2]

Respondents' inability or unwillingness to provide accurate data will not be a problem, though, if an appropriate observation method is used. For instance, management can accurately assess the sales impact of the special display by programming the cash registers to observe sales

---

[2] Catherine L. Harris and Deborah C. Wise, "It's Getting Harder for Computer Gurus to Sound Wise," *Business Week*, May 27, 1985, p. 108.

(method F) in an experimental group of stores having the special display and in a comparable control group of stores not having the display.

In summary, when the *same type* of data (e.g., sales of Kwality brand shirts or number of customers who saw the special display for those shirts) can be obtained through questioning and through observation, the latter will generally yield more accurate data. Since respondents do not directly interact with researchers in the observation approach, data distortions stemming from respondents are minimized. In other words, to the extent that the observation approach does not involve verbal or written answers from respondents, it is likely to produce more accurate data than the questioning approach.

Bear in mind, however, that the subjectivity or the carelessness of an observer may lead to inaccuracies in observational studies. For instance, in method D the observer may see a customer viewing the display for several minutes and infer that the customer is impressed by the display, when in actuality the customer may be frowning at the display out of annoyance. Just as interviewers may bias data collected through questioning, observers may cause inaccuracies in observational data. Furthermore, as was pointed out earlier, observation cannot provide accurate data on perceptions, motives, attitudes, or other inner feelings of people.

Sorunu

Respondent Convenience

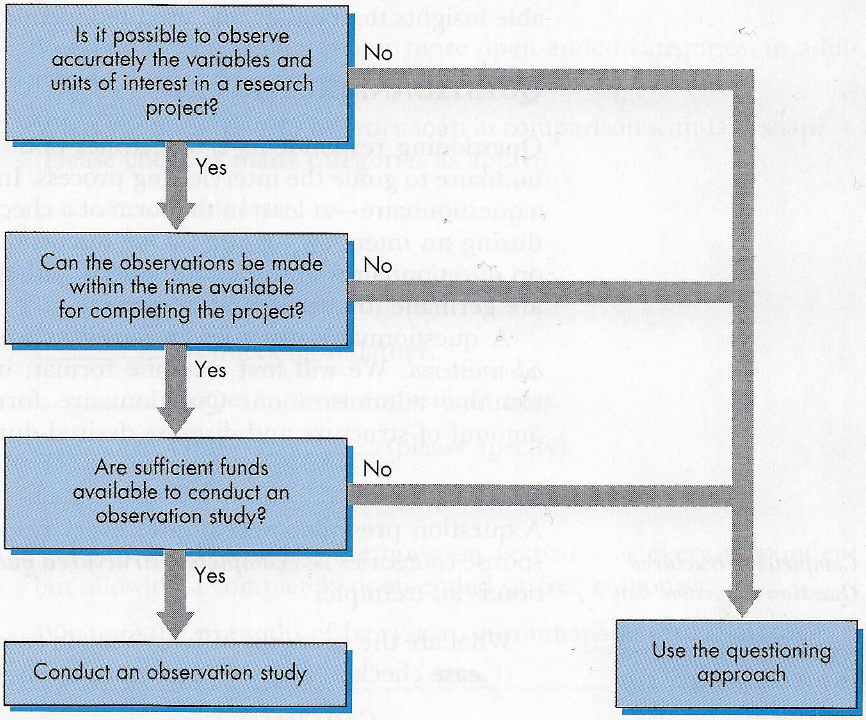
By and large, observation studies will be more convenient from the respondent's standpoint and hence are superior in terms of gaining the respondent's cooperation during data collection. This conclusion follows from the fact that respondents do not actively participate in the data collection in observation studies. The few exceptions where observation studies may be somewhat inconvenient to respondents involve situations in which they have to go to a particular place at a particular time to participate in a study (more will be said about such studies later).

In summary, a major limitation of the observation method is its inability to generate data about variables that are not visible. The questioning method is more versatile in this regard and, hence, is more widely used in practice.[3] Another drawback of the observation method is that an observer will typically have to wait for relevant events or behavior to occur. Therefore, it is likely to involve more time and expense, although modern advances in observation methodologies are starting to mitigate these deficiencies. A key strength of observation studies is that they are

---

[3] Since questionnaire studies apparently dominate the marketing research field, researchers may not even consider observation as an option. For a discussion highlighting the potential capabilities of observation and emphasizing the overuse of questionnaires, see Michael L. Ray, *Unobtrusive Marketing Research Techniques* (Cambridge, Mass.: Marketing Science Institute, 1973).

Exhibit 6.1 *Choosing Between the Questioning and Observation Approaches*


Is it possible to observe accurately the variables and units of interest in a research project?
No
Yes
Can the observations be made within the time available for completing the project?
No
Yes
Are sufficient funds available to conduct an observation study?
No
Yes
Conduct an observation study
Use the questioning approach

likely to provide more accurate data since distortions stemming from respondents will be much lower than in studies employing the questioning approach. Moreover, the absence of direct respondent participation in the data collection makes the observation method more convenient to respondents and minimizes the need to secure their cooperation—a significant hurdle in most studies involving questioning of respondents. Therefore observation appears to be the preferable method if the required data can be obtained through either of the two basic methodologies. Exhibit 6.1 lays out a simple framework to help in choosing between them.

In some research settings, however, a combination of the two methods can be used. For instance, in method D (involving human observation) in the Kwality Knitware case, the observer can first unobtrusively record a customer's reactions to the display. Immediately after the customer leaves the display area, the observer can conduct a personal interview with him or her to obtain additional data and to verify the observer's interpretation of the customer's reactions. Of course, well-trained observer-interviewers, as well as adequate time and monetary resources, will be necessary for such an approach to be beneficial. Resources permitting, the observation-questioning combination may yield more valuable insights than either one used independently.[4]

QUESTIONNAIRE FORMAT

Questioning respondents is sometimes done informally, with no questionnaire to guide the interviewing process. In many instances, however, a questionnaire—at least in the form of a checklist of items to be covered during an interview—is used. Our discussion in this section will focus on questionnaires and their variations, although many of the concepts are germane to questioning in general.

A questionnaire can vary in *format* as well as in terms of *how it is administered*. We will first examine format; in the next section we will examine administration. Questionnaire format is a function of the amount of structure and disguise desired during data collection.

Structure

Completely Structured Question A question that is presented verbatim to every respondent and has fixed response categories.

A question presented verbatim to every respondent and with fixed response categories is a ***completely structured question***. The following question is an example:

What are the strengths of Ivory soap in comparison with Dial soap? (Please check as many categories as apply.)

_________ Costs less.

_________ Lasts longer.

_________ Smells better.

_________ Produces more lather.

_________ Comes in more convenient sizes.

Completely Nonstructured Question A question that is not necessarily presented in exactly the same wording to every respondent and does not have fixed responses.

At the other extreme is the ***completely nonstructured question***, one that is not necessarily presented in exactly the same wording to every respondent and does not have fixed responses. For instance, instructing an interviewer to "probe a respondent's perceptions of Ivory's strengths relative to Dial" is tantamount to using a completely nonstructured question—one that is very flexible from the standpoint of both the interviewer and the respondent.

---

[4] For further elaboration of the importance and benefits of using multiple data collection methods, see Ray, *Unobtrusive Marketing;* see also Eugene J. Webb, Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally, 1966), Chap. 7.

Questions characterized by varying degrees of structure between the two extremes are also possible. Thus more flexible versions of the completely structured question presented earlier can be designed, as illustrated next.

Version 1: Providing one or more open-ended categories, in addition to the fixed response categories already included:

What are the strengths of Ivory soap in comparison with Dial soap? (Please check as many categories as apply.)

_________ Costs less.

_________ Lasts longer.

_________ Smells better.

_________ Produces more lather.

_________ Comes in more convenient sizes.

_________ Other _________ (please specify).

_________ Other _________ (please specify).

Version 2. Presenting the question verbatim to every respondent but allowing a completely open-ended or free response:

What are the strengths of Ivory soap in comparison with Dial soap?

____

____

____

____

In questionnaire studies these two versions are more frequently used than the two extremes. Moreover, in practice, a typical structured question is similar to version 1, and a typical nonstructured question (also known as an ***open-ended question***) is similar to version 2.

Open-Ended Question
A question that respondents are free to answer in their own words.

What are the relative advantages and drawbacks of structuring questions? We can compare structured and nonstructured questions on the same dimensions we used to compare the questioning and observation methods—namely, versatility, time, cost, accuracy, and respondent convenience. Table 6.1 provides such a comparison.

Structured questions have an overall edge over nonstructured questions with respect to speed and cost of data collection and analysis, and convenience to respondents. However, the inflexibility of structured questions makes them inappropriate for studies whose primary purpose is to generate new ideas (i.e., exploratory research). Moreover, the accuracy or objectivity of data collected through structured questions will not

Table 6.1 *Advantages and Disadvantages of Structured Questions Relative to Nonstructured Questions*

Evaluation Criteria	Advantages of Structured Questions	Disadvantages of Structured Questions
Versatility	Can be used to study diverse populations; literacy levels and communication skills of respondents not as critical as for nonstructured questions More topics/issues can be covered in interview/questionnaire of given length	Not as good in providing new insights/ideas as nonstructured questions Cannot obtain in-depth or detailed responses
Time	Less time to respond as well as to record responses Collected data can be quickly transferred to computer memory for analysis; in some studies recording coded responses directly into computer terminals as interview is taking place may be possible	May take more time to design, unless researcher has clear idea of what to ask and what specific responses to expect
Cost	Cheaper since interviewer time and skill levels required to record and interpret data are usually lower than for nonstructured questions	
Accuracy	Less chance of interviewer and respondent errors in recording answers	No guarantee that checked responses fully and/or truly reflect respondents' intended answers
Respondent convenience	More convenient to respondents in terms of time needed to respond and ease of responding	

be high unless all potential answers to a question are unambiguously implied by the precoded response categories.

The limitations of structured questions offer a clue about the circumstances in which they will be appropriate. Structured questions are likely to be most suitable in a situation where (1) uncovering new ideas is *not* the main objective and (2) the researcher has a good feel for the range and types of responses so that meaningful and valid response categories can be constructed. A research setting likely to possess these two features is one where the research purpose and data requirements are clear, from preliminary research results or relevant past experience, and hence the need is for research of a conclusive nature. This inference is also consistent with our discussion in Chapter 4 concerning differences in the nature of data collection in exploratory- and conclusive-research projects (see Table 4.1 in Chapter 4).

In short, the degree of finality of a research project will have a direct bearing on the extent of questionnaire structure. The more conclusive the research, the more structured the questionnaire can be. Likewise, the more exploratory the research, the more nonstructured the questionnaire can, or perhaps should, be. Nonstructured questionnaires are ideal for exploratory-research projects not only because they have insight-generating potential but also because the time and cost drawbacks are mitigated somewhat by the small sample sizes and, hence, the low volume of collected data.

We now turn to the second determinant of questionnaire format, namely, disguise.

Disguise

Disguised Question
An indirect question whose true purpose is not obvious to respondents.

Nondisguised Question
A direct question whose true purpose is obvious to respondents.

A ***disguised question*** is an indirect question whose true purpose is not obvious to respondents. Disguised questions are used to examine issues for which direct questions (i.e., ***nondisguised questions***) may not elicit truthful answers. Using direct questions to investigate sensitive topics or topics that may be embarrassing to respondents will most likely yield data of questionable accuracy. Consider the following direct, or nondisguised, question:

> Would you buy a cheap brand of liquor and serve it from a bottle from an expensive brand to impress your guests?
>
> ______ Yes ______ No

As you might expect, respondents who practice the deception described by this question will prefer not to answer it.[5] But since refusing to answer may be interpreted as a yes response, they will invariably be tempted to answer no. Indeed, to obtain valid data about the number of respondents who engage in the practice described in the question may be impossible.

However, a useful surrogate measure of respondents' feelings toward the practice can be obtained by asking a disguised question like the following one:

---

[5] Further discussion of this topic can be found, for example, in Claire Selltiz, Lawrence S. Wrightsman, and Stuart W. Cook, *Research Methods in Social Relations* (New York: Holt, Rinehart and Winston, 1976).

Would people you associate with buy a cheap brand of liquor and serve it from a bottle from an expensive brand to impress their guests?

_________ Yes _________ No

This question may be more acceptable to respondents than the direct question, and answers to it may provide a more accurate measure of respondents' own inclinations toward serving cheap liquor out of bottles from expensive brands.

Respondents may sometimes be *unable* to provide accurate responses to direct questions, even if they are willing to answer such questions. Examples of situations in which this difficulty is likely to arise include studies of personality traits and purchase motives wherein the desired information may exist at the respondents' subconscious or unconscious level. Disguised questions will be more appropriate than direct questions under those circumstances.

In general, disguised questions are preferable when respondents are likely to be unwilling and/or unable to answer direct questions accurately. However, constructing suitable disguised questions and properly interpreting the responses require special skills and training in such areas as psychology and psychoanalysis. Therefore using disguised questions can become quite expensive if it is to be done correctly.

Types of Questionnaires

The degree of finality of a research project and the extent to which respondents are willing and able to answer direct questions about the issues to be examined determine the types of questions to be used. From these two characteristics, four broad categories of questionnaire formats can be identified, as shown in Table 6.2. They are structured, nondisguised; nonstructured, nondisguised; nonstructured, disguised; and structured, disguised.

Since our discussion thus far has focused on characteristics of *questions* and we are now shifting our focus to types of *questionnaires,* you should note that structured questionnaires, for the most part, consist of a set of

Table 6.2. *Determinants of Questionnaire Format*

Degree of Finality of Research	**Respondent Willingness/Ability to Answer Direct Questions**	
	High	**Low**
High	Structured, nondisguised questionnaire	Structured, disguised questionnaire
Low	Nonstructured, nondisguised questionnaire	Nonstructured, disguised questionnaire

structured questions that are presented verbatim, and typically in the same order, to all respondents. Nonstructured questionnaires, for the most part, consist of a set of open-ended questions that may not necessarily be presented verbatim, and in the same order, to all respondents. Likewise, disguised questionnaires, for the most part, consist of a set of questions wherein the purpose of each question is not obvious to respondents. In contrast, nondisguised questionnaires, for the most part, consist of a set of direct questions whose purpose should be obvious to respondents. Each of the four types of questionnaire formats shown in Table 6.2 is discussed next.

Structured, Nondisguised Questionnaires. Structured, nondisguised questionnaires are widely used in marketing research studies, especially those involving large sample sizes. Structured, nondisguised questionnaires have all the advantages of structuring we saw earlier. They are especially appropriate for descriptive-research studies where the issues to be examined are clear-cut and there is no need for any disguise.[6]

nekvia

Research in Use 6.3 contains a sample section from a structured, nondisguised questionnaire used by the Bankers Life and Casualty Company ("Bankers") in a mail survey of new insurance policy owners. Several aspects of the partial questionnaire in Research in Use 6.3 illustrate key points made earlier. First, the questions have prespecified response categories. However, notice that question 4 and the "Other" category in question 7 are open-ended; but since they can be answered with very few words, the questionnaire is essentially structured. In fact, it is not unusual for a typical structured questionnaire to have a few open-ended items. Second, since the questions and the response categories are straightforward, the questionnaire is likely to be quite convenient and easy to answer for a wide variety of respondents. Third, preparing the collected data for analysis will also be quite easy and fast. Notice the numbers printed beside each response category. These numbers facilitate rapid coding of the responses after the questionnaires are returned. Numbers within parentheses refer to the computer record positions in which responses to the corresponding questions are to be stored; numbers without parentheses are the code numbers corresponding to the various responses.

Nonstructured, Nondisguised Questionnaires. The flexibility and the directness of nonstructured, nondisguised questionnaires make them ideally suited for situations in which a researcher wants to give respondents a free hand in providing information. Hence the nonstructured,

---

[6] For an elaboration on the pros and cons of structured, nondisguised questionnaires, see Selltiz, Wrightsman, and Cook, *Research Methods in Social Relations.*

Research in Use 6.3 *Bankers Life and Casualty Company*

Sample Section from a Structured, Nondisguised Questionnaire

3. Did your new Bankers policy replace a policy you had before? (11)

Yes ☐ 1   No ☐ 2

4. From what company did you purchase your old policy? ____________ (12-13)

I did not have this coverage before. ☐ 99

5. How long ago did you purchase your old policy? (14)

Within the last two years .... ☐ 1
Two to five years ago .... ☐ 2
Over five years ago .... ☐ 3
I did not have this coverage before .... ☐ 4

6. Had you heard of Bankers before buying your new policy? (15)

Yes .... ☐ 1
No .... ☐ 2
Don't recall .... ☐ 3

7. Based on your knowledge of Bankers, by reputation or personal experience, how would you rate it in each of these areas:

	Above Average 1	Average 2	Below Average 3	Don't Know 4	
Financially sound	☐	☐	☐	☐	(16)
A large, well-known company	☐	☐	☐	☐	(17)
Offers good policies at reasonable cost	☐	☐	☐	☐	(18)
Has competent, well trained agents	☐	☐	☐	☐	(19)
Answers inquiries promptly and fully	☐	☐	☐	☐	(20)
Pays claims promptly and fairly	☐	☐	☐	☐	(21)
Serves the needs of senior citizens	☐	☐	☐	☐	(22)
Serves the needs of families	☐	☐	☐	☐	(23)
Other ____________	☐	☐	☐	☐	(24)

Source: Marketing Research Department, Bankers Life and Casualty Company, 1000 Sunset Ridge Road, Northbrook, Ill. 60062. Used by permission.

nondisguised format is very popular in exploratory-research projects involving personal interviews. In many such instances the questionnaire merely consists of a checklist of relevant issues to be covered, and the interviewer is free to change the wording as well as sequencing of the issues to make the interview flow smoothly and naturally. The main idea is to let respondents provide as much information, in as unrestricted a fashion, as possible. Interviews of this nature are appropriately called ***depth*** (or sometimes ***in-depth***) ***interviews***.

Depth (In-Depth) Interview *A form of nonstructured, nondisguised questioning in which respondents are encouraged to provide as much information, in as unrestricted a fashion, as possible.*

Depth interviews may be more useful than structured interviews in industrial marketing research surveys.[7] According to J. M. Smith, in these surveys,

> the respondents, unlike those in consumer market research, are usually experts holding senior managerial posts in business or industry. Such people do not react at all favorably to the rigidity of a formal, structured interview. It has been found from experience that most expert respondents have a story to unfold about some aspect of the subject of the survey. This "story" is usually very valuable and it should not be suppressed because of the limits imposed by the precoded question.[8]

For example, recall the industrial marketing research project conducted for the Wayne Engineering Corporation (the garbage packer manufacturer) that was described in Chapter 3. Personal interviews in this project were conducted without the aid of a formal questionnaire; the consultant only had a list of issues to be discussed during each interview. After raising a particular issue, the consultant simply listened and took notes while the respondent did most of the talking.

A variation of the depth interview is the focus group interview, involving the simultaneous interviewing of a group of respondents. Several examples of research projects using focus group interviews were presented in earlier chapters. Focus group interviewing is so widely used that a detailed discussion of it is presented in the next chapter. We will therefore not elaborate on it here.

Motivation Research
Research intended to uncover inner feelings and motives that people are unable or unwilling to reveal when questioned directly.

Nonstructured, Disguised Questionnaires. Nonstructured, disguised questionnaires are primarily used in so-called ***motivation research***. "Motivation research is a phase of marketing research which attempts to answer the question, 'Why?' . . . Motivation research seeks to relate behavior to underlying processes such as people's desires, emotions, and intentions."[9] Consumers are at times unable or unwilling to reveal their true reasons for buying a particular product or reacting in a certain way to a marketing stimulus. That is, consumers may consciously or unconsciously try to suppress their motives.

[illegible]
haluja tunteita
aikomuksia

For instance, a socially upscale man who is motivated to buy his clothes from a discount store because of cost considerations will be very reluctant to discuss his purchases. A woman who does all her cooking from scratch because she feels "good" women should do so may be unwilling to reveal her true motives for fear of being labeled "old-fashioned." In situations like these a direct-questioning approach is likely to lead to either no responses or less-than-truthful responses. What is needed is a procedure in which a person can feel uninhibited enough to describe freely (i.e., in a nonstructured fashion) his or her inner feelings, by projecting them onto a setting that supposedly is unrelated to the respondent (i.e., disguised). Such nonstructured, disguised questioning procedures are called [illegible] ***projective techniques***.

---

[7] William E. Cox, Jr., *Industrial Marketing Research* (New York: Wiley, 1979), p. 245.

[8] Joan Macfarlane Smith, *Interviewing in Market and Social Research* (London: Routledge & Kegan Paul, 1972), p. 142.

[9] George Horsley Smith, *Motivation Research in Advertising and Marketing* (New York: McGraw-Hill, 1954), p. 3.

Projective Technique
A nonstructured, disguised form of questioning in which [illegible] *a fairly ambiguous stimulus is presented to respondents, who are then asked to react to or describe the stimulus.*

Although different types of projective techniques are available, they all share two common features: (1) A fairly ambiguous stimulus is presented to respondents, and (2) in reacting to or describing the stimulus, the respondents will indirectly reveal their own inner feelings.[10] Examples of projective techniques used in marketing research include the word association test, the sentence completion test, the thematic apperception test (commonly referred to as TAT), and the cartoon test. We will briefly examine each of these.

Word Association Test
A nonstructured, disguised form of questioning in which respondents are asked to express the first word that comes to mind in response to each word on a list of words. In a typical word association test the words are read aloud, one at a time, to each respondent.

The stimulus in ***word association tests*** is a list containing anywhere from just a few to over a hundred words of interest mixed in with somewhat irrelevant or neutral words designed to preserve the disguised nature of the study. For instance, a list in a study dealing with home computers may include relevant words such as prestige, bookkeeping, and video games as well as neutral words such as cooking, exercise, furniture, and newspapers. Word association tests can be especially useful in uncovering people's feelings about new products or services, brand names (recall from Chapter 1 that Ford Motor Company used word association tests in evaluating the name Edsel), and key words being considered for use in advertising copy or other promotional materials.

In a typical word association test the words are read aloud, one at a time, to each respondent. The respondent is asked to say the first word that comes to mind as soon as each stimulus word is presented. Interpretation of the responses is not easy. It is based on several factors and includes analyzing the meanings of the response words, the overall nature of responses given by each respondent, the patterns of responses across respondents, the time taken to respond to each stimulus word, and the physical reactions of respondents.[11]

Sentence Completion Test
A nonstructured, disguised form of questioning in which respondents are asked to finish a set of incomplete sentences. A typical sentence completion test is given to a group of respondents who are asked to finish the sentences in writing.

A ***sentence completion test*** involves asking respondents to finish a set of incomplete sentences. Each incomplete sentence consists of a few words related to or neutral to the topic being researched, followed by a blank space for the response. To illustrate, a study dealing with people's inner feelings toward "Buying American" may contain the following sentences:

---

[10] The concept of projection, which is the key to any projective technique, is actually quite complex and has been given differing shades of meanings in the psychology literature; for a discussion of this issue, see Boris Semeonoff, *Projective Techniques* (New York: Wiley, 1976), Chap. 2.

[11] Detailed discussion of the analysis and the interpretation of responses to word association tests, as well as to other projective techniques, is beyond the scope of this textbook. However, several books on motivation research that treat this topic in detail are available; see, for example, Smith, *Motivation Research.*

American automobiles ______________________.

Restrictions on imports ______________________.

Every U.S. citizen ______________________.

Foreign-made products ______________________.

Unemployment in the United States ______________________.

Sentence completion tests are typically given to groups of respondents who are asked to finish the sentences in writing. Neither the time taken to respond to each item nor the respondents' physical reactions are monitored, as they are in word association tests. Hence sentence completion tests are easier to administer.[12] Moreover, they allow researchers greater flexibility to give direction to the stimuli so that the responses obtained are more likely to be associated with the subject matter under study. In contrast, word association tests, by pressuring respondents to react quickly to single words, may lead to answers that are hard to interpret. However, there is a price for the greater flexibility and ease of administration of sentence completion tests: The respondents can think through their answers, and therefore the information obtained is limited to what each respondent is *willing* to divulge.

Thematic Apperception Test (TAT)
A nonstructured, disguised form of questioning in which respondents are shown a series of pictures, one at a time, and are asked to write a story about each.

The ***thematic apperception test*** (TAT) was first developed by Henry A. Murray to measure personality.[13] The original TAT consisted of a set of 20 cards with pictures printed on them. However, adaptations of the TAT have since been developed for special groups such as children.[14] In a typical administration of the TAT, the respondent is shown each picture for a short period (about 20 seconds) and is asked to write a story about it in 20 minutes or so. The respondent is specifically asked to describe such things as what is happening in the picture, why it is happening, and what the feelings of the characters in the picture are. These stories are interpreted by specially trained analysts to ascertain the respondent's personality. The TAT can be administered individually or in groups.[15]

---

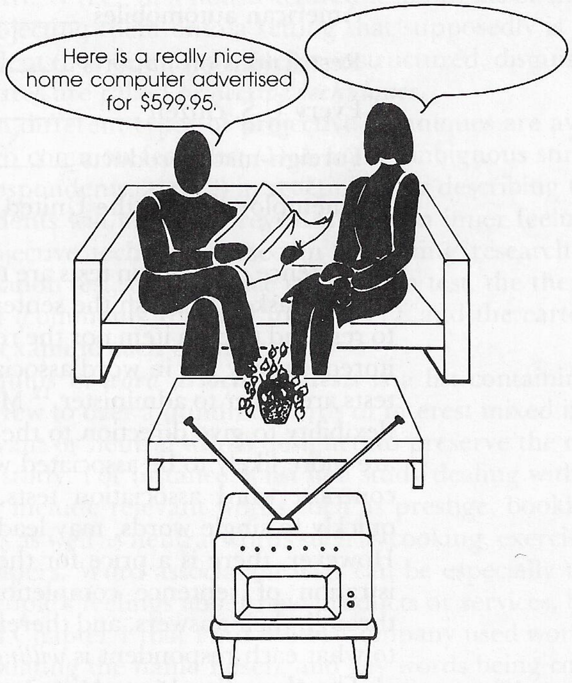
[12] For further details about the comparison between administration of word association tests and administration of sentence completion tests, see Paul G. Datson, "Word Associations and Sentence Completion Techniques," in *Projective Techniques in Personality Assessment,* ed. A. I. Rabin (New York: Springer, 1968), pp. 264–289.

[13] Henry A. Murray, *Explorations in Personality* (New York: Oxford University Press, 1938).

[14] Charles Neuringer, "A Variety of Thematic Methods," in *Projective Techniques in Personality Assessment,* ed. A. I. Rabin (New York: Springer, 1968), pp. 222–261.

[15] For further details concerning TAT administration, see, for example, Smith, *Motivation Research,* pp. 126–130.

Exhibit 6.2 *Cartoon test*


Here is a really nice home computer advertised for $599.95.

Adaptations of the TAT have been employed in marketing research by using pictures specifically designed for the research topic. To illustrate, in a study of attitudes toward magazine reading, "a picture was included which showed a family sitting in a living room reading. In making up stories for this picture, respondents revealed many of their views about magazines, the family and home, what the husband reads, what the wife reads, and so on."[16] Other possible applications of TAT techniques in marketing research include the evaluation of pictures being considered for use in advertisements, in promotional brochures, and in product packaging.

Pinnos

Cartoon (Balloon) Test
A nonstructured, disguised form of questioning in which respondents are shown one or more cartoon pictures with empty "balloons" and are asked to fill the balloons with words reflecting thoughts or verbal statements of the characters involved.

A ***cartoon test*** (sometimes also known as a ***balloon test***) is another pictorial technique like the TAT. The stimuli in cartoon tests are line drawings such as the one shown in Exhibit 6.2. The cartoon in Exhibit 6.2 can, for instance, be used to gain insight into housewives' attitudes toward home computers.

---

[16] Smith, *Motivation Research*, pp. 142–143.

The respondent in a cartoon test is asked to examine the stimulus picture and fill in the empty "balloons" with words reflecting thoughts or verbal statements of the characters involved. Unlike the stories required in a TAT, responses to a cartoon test are limited and specific. Therefore the time needed to respond as well as the volume of responses to be analyzed should be less. Cartoon tests have been used in a number of marketing research projects, including one conducted by State Farm to determine customer perceptions about automobile insurance promoted with a price appeal.[17]

The tests described so far by no means exhaust the variety of projective techniques that can be, and are being, used in marketing research. For instance, Child Research Services (CRS), a division of McCollum/Spielman Associates, Inc., employs a number of projective techniques in researching children. According to CRS:

> A variety of projective children's games are used in service of a particular client's needs. These include charades, sentence completions and word associations, drawing or completing pictures, creating dialogue for "bubbles" above characters in cartoon-style pictures. There's also the whispering of "secret votes" to the group leader, and children are frequently asked to "pretend telephone," projecting themselves into the role of parents describing a recent purchase.
>
> Youngsters may also be asked to "act out" commercials in a wide variety of shopping and domestic situations. Out of these creative play sessions, many a commercial has been re-structured for more effective communication to children; prototypes for many new products have been designed based on children's uses and needs; and many products have been revitalized to enhance their appeal to children.[18]

A researcher's imagination is the only limit on the uniqueness of a projective technique to be used in a situation. Indeed, one of the first applications of motivation research in a marketing context—one that is cited as a classic example in many textbooks—involved the use of a very creative and interesting projective technique. The study was conducted in response to unexpected customer resistance to Nescafe instant coffee when it was first introduced some 40 years ago.[19] When women were questioned directly about why they did not like instant coffee, the typical answer was that they did not like its flavor. However, suspecting that consumers were merely using flavor as an excuse, the researchers devised a unique projective technique to get at the real reason(s) for the resistance to instant coffee.

---

[17] Joseph W. Newman, *Motivation Research and Marketing Management* (Boston: Harvard University, Division of Research, 1957), pp. 80–81.

[18] "The Child Research Services," a brochure prepared by Child Research Services, a division of McCollum/Spielman Associates, Inc., 235 Great Neck Road, Great Neck, N.Y. 11021.

[19] Mason Haire, "Projective Techniques in Marketing Research," *Journal of Marketing*, 14 (April 1950), pp. 649–652.

The researchers selected two comparable samples of 50 housewives each. They also drew up the following two shopping lists:

List 1	**List 2**
Pound and a half of hamburger	Pound and a half of hamburger
2 loaves of Wonder Bread	2 loaves of Wonder Bread
Bunch of carrots	Bunch of carrots
1 can Rumford's baking powder	1 can Rumford's baking powder
Nescafe instant coffee	Maxwell House coffee (drip ground)
2 cans Del Monte peaches	2 cans Del Monte peaches
5 lb potatoes	5 lb potatoes

The women in one sample were shown list 1, and those in the other sample were shown list 2. Each respondent was asked to examine the shopping list and then write a description of the type of woman who could have made out the list.

The descriptions of the hypothetical shopper that emerged from the two samples were remarkably different. The sample that saw list 1 tended to describe the woman as a lazy person, a poor planner, and a spendthrift. In contrast, the descriptions provided by the respondents who saw list 2 were much more favorable to the hypothetical shopper: Many of them perceived her as a good housewife who was thrifty. These findings implied that the convenience feature of instant coffee, which was being used as a unique selling point for Nescafe, was apparently backfiring owing to its incongruence with society's expectations of homemakers. As a result of this study, the promotional theme for Nescafe was modified by toning down the emphasis on convenience.

In short, the nonstructured, disguised, or projective, method of data collection is flexible enough to be tailored to fit any research setting calling for an examination of people's inner feelings. Projective techniques can be valuable if properly used. Unfortunately, designing valid projective tests and correctly interpreting the results of these tests is not easy. Very special skills and training are required. Researchers who are fascinated by projective techniques but are not qualified to use them can end up misusing them. Thus, after being widely used in the 1950s to solve marketing problems, projective techniques have been gradually losing their popularity, primarily because of inappropriate use by unqualified researchers who could not deliver on their promises to decision makers. Today there is considerable skepticism, especially among practitioners, about the whole area of motivation research. Even the validity of the shopping list study just described has been challenged. A senior research psychologist working for du Pont has presented evidence showing that items in the lists other than coffee could have biased respondents' perceptions about the hypothetical shopper.[20]

salatti

Structured, Disguised Questionnaires. Structured, disguised questionnaires are generally used to uncover people's attitudes toward sensitive issues of concern to society, like abortion, pollution, or deregulation. A structured, disguised questionnaire consists of a number of factual items to which respondents provide structured answers such as yes or no and true or false. A wide variety of items—ranging in degree of favorableness toward the issues being investigated—are included in the questionnaire. The items themselves can be real or fictitious.

The rationale behind a structured, disguised test is the following assumption: What and how much people *claim they know* about an issue can shed light on their attitudes toward the issue. For instance, a person who favors deregulation would tend to agree with, or label as correct, many more supposedly factual statements that are proderegulation than those that are antideregulation. So that respondents indirectly expose their inner feelings, they are deliberately told that each item has a correct and an incorrect response: "In contrast with most projective tests, in which the individual is encouraged to believe that there are no objectively 'correct' answers, the subject [in a structured, disguised test] is led to believe that there are 'right' and 'wrong' responses, and the attempt is made to motivate him to do as well as possible on the test."[21]

Structured, disguised tests, while somewhat more convenient to answer and code than nonstructured, disguised tests, are not any easier to design and interpret. For this reason, and perhaps also because these tests are typically designed to gauge people's attitudes toward fairly broad issues, they are not widely used in marketing research.

Thus far we have examined a variety of questionnaire methods classified according to their format. Let us now look at different methods of administering questionnaires.

QUESTIONNAIRE ADMINISTRATION METHODS

Personal Interview
A questionnaire administration method in which the interviewer and respondent have face-to-face contact.

In the Kwality Knitware example the first three methods (A, B, and C) describe the basic approaches for collecting data through questionnaires:

> ***Personal interviews***, in which there is face-to-face contact between interviewers and respondents (method A).

---

[20] James C. Anderson, "The Validity of Haire's Shopping List Projective Technique," *Journal of Marketing Research,* 15 (November 1978), pp. 644–649.

[21] Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook, *Research Methods in Social Relations,* rev. ed., 1 vol. (New York: Henry Holt, 1959), p. 300.

Table 6.3 *Changes in Popularity of Questionnaire Administration Methods*

	Percent of Largest U.S. Consumer Goods and Services Companies Using Method		
Method	**1978**	**1983**	**1987**
Personal Interviews			
In-home	61	47	39
Shopping mall intercept	89	90	86
Telephone Surveys			
Central WATS	90	91	98
Local telephone	67	61	*
Mail Surveys			
Surveys of panel households	53	57	67
Surveys of households at large	46	33	43

Source: Adapted from "Practices, Trends and Expectations for the Market Research Industry 1987," a report published by Market Facts, Inc., 676 North St. Clair Street, Chicago, Ill. 60611, April 29, 1987, p. 23. Used by permission.
*Not measured in 1987 survey.

Telephone Survey *A questionnaire administration method in which there is only voice contact between the interviewer and respondent.*

Mail Survey *A questionnaire administration method in which there are no interviewers and no direct contact between the researcher and respondent.*

Shopping Mall Intercept Interview *A form of personal interviewing in which respondents are intercepted in shopping malls and interviewed.*

Telephone surveys, in which there is only voice contact between interviewers and respondents (method B).

Mail surveys, in which there are no interviewers (method C).

Personal interviews have traditionally been conducted "door to door" or "in-home." However, this approach is declining in popularity owing to increasing difficulty in finding adult respondents at home during hours that are convenient for in-home interviewing, and getting respondents to cooperate if they are home. Most personal interviews these days are conducted in shopping centers or malls. In these interviews, called ***shopping mall intercept interviews***, respondents are intercepted in shopping malls and interviewed. Evidence of the shift in the mode of personal interviewing is revealed by Table 6.3, which summarizes findings from a series of studies (conducted by Market Facts, Inc., a leading research firm) about the survey methods used by some of the largest marketers of products and services to consumers. Table 6.3 also indicates growing usage of mail surveys and telephone interviews conducted from central locations through WATS (Wide Area Telephone Service) lines.

As demonstrated by the Kwality Knitware illustration, any of the three basic questionnaire administration methods can generally be used in a research project requiring primary-data collection. But not all of them may be equally appropriate in a given situation. While each method has certain advantages and limitations, the strength of the advantages and the seriousness of the limitations depend on the specific features of the situation. Keep this point in mind throughout the following discussion of the advantages and disadvantages of the three methods on the dimensions of versatility, time, cost, accuracy, and respondent convenience.

Versatility

A number of factors affect the versatility of questionnaire administration methods. Examples include the amount of data that can be collected, the types of questions that can be asked, and the kind of stimuli that can be presented to respondents during data collection. As you might expect, personal interviews are much more flexible and allow for the collection of a greater variety of data than telephone or mail surveys. Thus personal interviews are better than the other approaches for obtaining customer reactions to the appearance of a package, the taste of alternative formulations of a food product, and so on. Telephone and mail surveys are somewhat deficient in terms of presenting stimuli to respondents, although modern technological advances are improving the versatility of these methods. For instance, one can present visual stimuli (including the questionnaires themselves) through respondents' television sets (via video cable) and obtain responses over the telephone or through other means of electronic feedback.[22]

The personal-interview method is also better when the research involves lengthy nonstructured questionnaires, since respondent resistance is likely to be especially high to such questionnaires if administered through telephone or mail surveys. In a face-to-face interview, once a respondent initially agrees to cooperate, it may be psychologically difficult for him or her to terminate the interview.

Telephone surveys are more flexible than mail surveys in the sense that there is at least verbal contact between the interviewer and respondent. This added flexibility of telephone interviews allows the interviewer to ask some open-ended questions and to probe a respondent as needed during the interview. Although mail questionnaires can generally have a larger number of questions than questionnaires administered over the telephone, they typically will have to be quite structured to encourage respondent cooperation, which, in turn, reduces flexibility.

In summary, with respect to overall versatility, personal interviews are perhaps the best, followed by telephone and mail surveys, in that order.

---

[22] One such interactive video data collection system, called QUBE, is already in place in Columbus, Ohio. For a description of this system, see "QUBE: The Ultimate Testing Device," *Marketing News,* May 2, 1980, p. 7.

Time

Telephone interviewing is usually the best form of questionnaire administration when information is needed in a hurry. Telephone surveys avoid the time-consuming travel involved in in-home personal-interview surveys (interviewer travel time, of course, is not a drawback in surveys using shopping mall intercepts). They also do not suffer from the postal delays and respondent tardiness generally present in mail surveys.

Computer-Assisted Telephone Interviewing (CATI)
A computerized telephone-interviewing system.

Certain ***computer-assisted telephone interviewing*** systems (also known as ***CATI*** systems), in addition to cutting down data collection time, can also reduce the time required for data analysis. Market Facts, Inc., is a firm having such a system. Exhibit 6.3 presents a schematic description of Market Facts' National Telephone Center in Chicago. Respondents in any geographic area of the United States can be interviewed through the 70 WATS interviewing stations at Market Facts' National Telephone Center. The questionnaires used in conducting interviews from this center appear on computer terminals (called cathode ray tubes, or CRTs) in front of the interviewers. The responses are entered directly into these terminals as the interviews take place. Therefore the collected data are immediately transferred to computer memory and can be readily processed at any time. For example, the firm can conduct interviews one evening and have the responses analyzed and ready for presentation the next day.[23]

Personal interviews can normally generate data faster than mail surveys, especially when adequate resources are available to hire as many interviewers as needed to collect the desired data within a specified time period. In contrast, the time needed to obtain data through mail surveys is essentially out of the hands of the researcher once the questionnaires are mailed out.

Thus on the time dimension the typical rank ordering of questionnaire administration methods is as follows: (1) telephone surveys, (2) personal interviews (in-home worse than telephone, but mall intercept quite comparable), and (3) mail surveys.

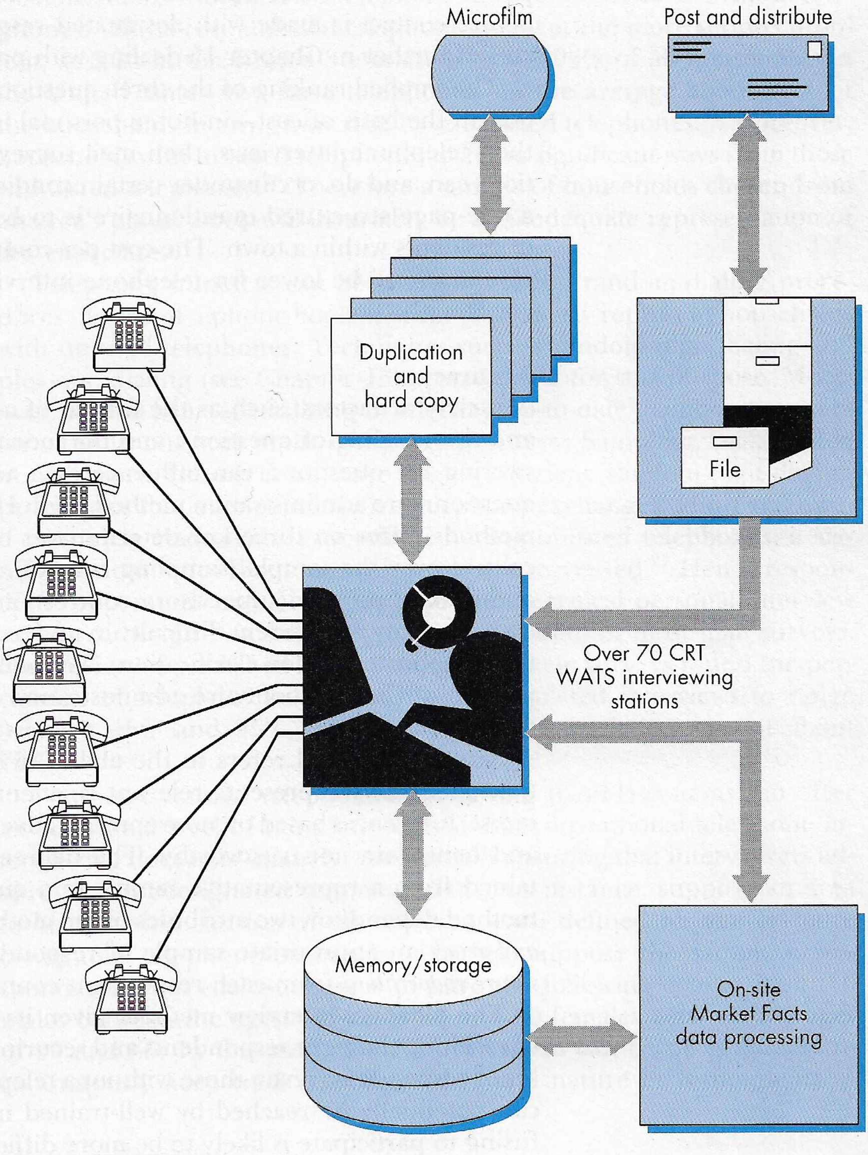
Cost

While mail surveys are generally the slowest for obtaining data, they are also the least expensive. They are the least expensive mainly because mail surveys have no interviewers, who usually account for a large share of the data collection costs in surveys using personal or telephone interviewing. In-home personal interviews are typically more expensive than telephone surveys because of travel expenses involved (travel expenses are not a major concern in shopping mall intercepts). Of course, telephone interviewing can also become quite expensive if lengthy long-

---

[23] Additional details about this telephone-interviewing system are presented in "Data Collection and Analysis for Reducing Business Decision Risks," a brochure prepared by Market Facts, Inc., 676 North St. Clair Street, Chicago, Ill. 60611.

Exhibit 6.3 *Market Facts' National Telephone Center*


Microfilm
Post and distribute
Duplication and hard copy
File
Over 70 CRT WATS interviewing stations
Memory/storage
On-site Market Facts data processing

Source: "Data Collection and Analysis for Reducing Business Decision Risks," a brochure prepared by Market Facts, Inc., 676 North St. Clair Street, Chicago, Ill. 60611, p. 12.

distance calls are needed. Nevertheless, even under those circumstances in-home personal interviews may be more expensive on a per-completed-interview basis, especially if one or several callbacks are necessary before contact is made with designated respondents (callbacks are discussed further in Chapter 15 dealing with potential errors in fieldwork).

The implied ranking of the three questionnaire administration methods on the basis of cost—in-home personal interviews (most expensive), then telephone interviews, then mail surveys—is not universal. Exceptions can, and do, occur under certain conditions. For instance, suppose a one-page structured questionnaire is to be administered to a sample of residents within a town. The cost per completed questionnaire in this case may well be lower for telephone interviewing than for a mail survey.[24]

Accuracy

A variety of factors, such as the format of a questionnaire, the content and the wording of questions, and the moods of respondents while answering the questions, can influence data accuracy, irrespective of the questionnaire administration method used. However, the administration methods differ on three key determinants of data accuracy: degree of control over the sample (sampling control), degree of control over the questioning process (supervisory control), and the ability to sense and overcome any respondent difficulty in understanding or answering questions (opportunity for clarification). We will therefore focus our comparison of the questionnaire administration methods along these three dimensions.

Sampling control refers to the ability to collect data from a sample that adequately represents relevant segments of the population of interest. Inferences based on nonrepresentative data cannot be generalized and hence are not trustworthy. The degree to which data can be obtained from a representative sample in a questionnaire administration method depends on two attributes of the method: (1) the ability to *identify and reach* an appropriate sample of respondents, and (2) the ability to *secure cooperation* from each respondent contacted.

The personal-interview method, given its versatility, is most capable of reaching the right respondents and securing cooperation. Any desired set of respondents (even those without a telephone or a mailing address) can potentially be reached by well-trained interviewers. Moreover, refusing to participate is likely to be more difficult for a respondent when face-to-face with an interviewer than when contacted by telephone or mail. Therefore cooperation rates are normally higher when personal interviews are used.

---

[24] Seymour Sudman, *Reducing the Cost of Surveys* (Chicago: Aldine, 1967), provides a discussion of various facets of evaluating survey costs, including cost versus data quality comparisons.

Telephone surveys face two potential hurdles in attempting to reach a representative sample of households within the general population: (1) households with unlisted telephones and (2) households without telephones. Of the two, unlisted telephones present the more serious problem, at least in the United States. While over 95% of all households in the United States now have telephones, on the average about 20% of the households in any given area have unlisted telephones.[25] Moreover, households with unlisted telephones differ in significant ways from those with listed telephones.[26] Therefore a sample of households chosen from an area's phone book will invariably be an inadequate representation of the entire area.

Telephone numbers can be chosen by using random-dialing procedures instead of a phone book in order to properly represent households with unlisted telephones. Techniques such as random-digit dialing and plus-one dialing (see Chapter 15) are available for this purpose. Moreover, several firms specializing in selecting up-to-date telephone samples for virtually any area of the country (e.g., Survey Sampling, Inc., located in Westport, Connecticut) have highly efficient random-digit-dialing procedures that eliminate inactive telephone exchanges. However, even if contact is made with respondents having unlisted telephones, a significant number of them may refuse to be interviewed.[27] Hence respondent cooperation may not be as high as in a typical personal-interview survey, although it is still likely to be better than in most mail surveys. One review of published studies using telephone surveys found the percentage of telephone calls resulting in completed interviews to range between 45% and 95%, implying a typical completion rate of about 70%.[28]

Computer-assisted telephone-interviewing (CATI) systems can offer an additional level of sampling control that conventional telephone interviewing does not—namely, *automatically* ensuring that interviewers adhere to a predesignated sampling plan. For instance, suppose an area to be surveyed consists of 10 subareas, each defined by one or more contiguous telephone exchanges. Further, suppose the sample of respondents from each subarea is to have the following composition: 50 males between the ages of 20 and 65 and 50 females between the ages of 20 and 65. In this scenario, a CATI system can be programmed to continuously keep track of the number and nature of respondents as interviews are completed, compare the profile of respondents with the desired profile, and disallow interviews with respondents who either do not fall into one of the desired demographic groups or fall into a group whose quota had already been filled. In this way, interviewer errors (both inadvertent and deliberate errors) in selecting respondents can be eliminated. This feature of CATI will be especially useful in studies involving complex sampling plans wherein the opportunity for interviewer errors is likely to be high.

---

[25] Tyzoon T. Tybjee, "Telephone Survey Methods: The State of the Art," *Journal of Marketing,* 43 (Summer 1979), pp. 68–78.

[26] Patricia E. Moberg, "Biases in Unlisted Phone Numbers," *Journal of Advertising Research,* 22 (August/September 1982), pp. 54–55; Gerald J. Glasser and Gale D. Metzger, "National Estimates of Nonlisted Telephone Households and Their Characteristics," *Journal of Marketing Research,* 12 (August 1975), pp. 359–361.

[27] Moberg, "Biases in Unlisted Phone Numbers."

[28] Tybjee, "Telephone Survey Methods."

Mail surveys are generally the worst in terms of the percentage of contacts resulting in completed questionnaires. Getting back over 50% of the questionnaires initially mailed is an exception. Moreover, studies have shown that people who fill out and return questionnaires may be significantly different from nonrespondents. For instance, people with higher incomes and education levels are more likely to respond to mail surveys than others.[29]

As you might expect, researchers have tried numerous respondent inducements to improve mail survey response rates, not all of which have been effective.[30] The more successful inducement techniques include mailing questionnaires by first-class mail, sending follow-up questionnaires, and enclosing incentives—monetary or nonmonetary—with the questionnaires. Using incentives has been found to be particularly effective in improving response rates, apparently since people receiving them may feel obligated to cooperate.[31]

Incentives offered to respondents, however, may increase the number of returned questionnaires without necessarily improving sample representativeness. That is, people who respond because of an incentive may differ on critical characteristics from those who do not respond at all. Therefore, although a mail survey can reach virtually any and every potential respondent, it is generally less likely to provide a representative final sample, and hence has less sampling control, than personal interviews or telephone surveys.

We do not mean to imply that the problem of nonrepresentative samples is limited to mail surveys. As we will see in Chapter 15, lack of responses from particular types of respondents that results in biased samples can occur in personal and telephone surveys as well. But this problem is generally more severe in mail surveys.

---

[29] Two review articles that present comprehensive discussions of mail survey response rates are Leslie Kanuk and Conrad Berenson, "Mail Surveys and Response Rates: A Literature Review," *Journal of Marketing Research,* 12 (November 1975), pp. 440–453; and Arnold S. Linsky, "Simulating Responses to Mailed Questionnaires: A Review," *Public Opinion Quarterly,* Spring 1975, pp. 82–101.

[30] For a discussion of these inducement techniques and studies using them, see Kanuk and Berenson, "Mail Surveys and Response Rates"; Linsky, "Stimulating Responses"; Don A. Dillman, *Mail and Telephone Surveys* (New York: Wiley, 1978).

[31] One study involving a mail survey of respondents in a sample of firms found that in terms of improving response rates, the *size* of the incentive really may not matter as much as that *some* incentive is included; see Milton M. Pressley and William L. Tullar, "A Factor Interactive Investigation of Mail Survey Response Rates from a Commercial Population," *Journal of Marketing Research,* 14 (February 1977), pp. 108–111.

The issue of supervisory control refers to the ability to minimize interviewer errors such as failure to follow instructions, mistakes in recording answers, and cheating.[32] In this regard mail surveys, since they involve no interviewers, have an edge over personal and telephone interviews (assuming, of course, that the mail questionnaire is well designed with clear instructions so that errors by the respondents when they interview themselves are avoided). The personal-interview method is especially prone to problems stemming from interviewer errors. Supervising the work of face-to-face interviewers is hard, since they typically work alone in the field, although certain validity checks can be made *after* the data collection (see Chapter 15). Even rigorous interviewer training cannot guarantee that fieldworkers will not make mistakes. Indeed, according to studies conducted by the U.S. Census Bureau: "The evidence is compelling that, even with very structured training programs, interviewers exhibit considerable variability in their data collection for certain items."[33]

Telephone interviews conducted from central locations (like Market Facts' National Telephone Center in Chicago discussed earlier) are better than personal interviews in terms of supervisory control over the data collection. Most such centers have special CRTs and audio equipment that supervisors can use to monitor any interviewer's work as the interviewing is taking place. Therefore interviewer mistakes can be quickly spotted and corrected. Moreover, the interviewers themselves are likely to be extra careful because they know that a supervisor can monitor their work at any time. Additionally, as pointed out earlier, telephone interviews conducted through a CATI system can be controlled to a degree by programming the system to prevent some forms of interviewer errors. Thus the system automatically offers a certain amount of "supervisory" control.

On the dimension of opportunity for clarification—that is, the ability to detect and overcome problems respondents may experience in answering certain questions—mail surveys are the worst since there is no direct interaction with the respondent. Hence errors are bound to occur in mail survey data if respondents misunderstand certain questions or are unsure of how to answer them. In contrast, personal and telephone interviews can reduce such errors by allowing respondents to seek clarification and enabling interviewers to provide the necessary clarification.

---

[32] A good discussion of interviewer-induced errors in survey research is available in Charles F. Cannell, Kent H. Marquis, and Andre Laurent, *A Summary of Studies of Interviewing Methodology,* U.S. Department of Health, Education, and Welfare, Publication no. (HRA) 77-1343, (Washington D.C.: U.S. Government Printing Office, 1977).

[33] Barbara Bailar, Leroy Bailey, and Joyce Stevens, "Measures of Interviewer Bias and Variance," *Journal of Marketing Research,* 14 (August 1977), p. 343.

Personal interviews are perhaps the best in this regard because face-to-face contact allows interviewers to visually detect any respondent confusion or difficulty that may occur even when respondents do not explicitly seek clarification.

In terms of overall accuracy—that is, taking into account sampling control, supervisory control, and opportunity for clarification—generalizing about which method is best and which is worst is difficult, since each has strengths and weaknesses along different dimensions pertaining to data accuracy. Therefore, which questionnaire administration method (or methods) will provide the most accurate data depends on the specific characteristics of a research setting.

Respondent Convenience

The mail survey method, which is least disruptive and most flexible from a participant's perspective, is naturally the best in terms of respondent convenience. Personal interviews are perhaps the worst, especially since respondents may find it difficult to say no when confronted by a persistent interviewer. Telephone surveys generally fall between the other two methods with respect to respondent convenience.

Deciding on Which Questionnaire Administration Method to Use

As with other decisions, the decision about questionnaire administration method depends on the research project. Table 6.4 provides a compar-

Table 6.4 *Comparison of Questionnaire Administration Methods*

	Ranking of Methods		
Criteria	**Best**	**Second Best**	**Third Best**
Versatility			
Number of questions	Personal	Mail	Telephone
Amount/variety of information	Personal	Telephone	Mail
Presentation of stimuli	Personal	Telephone	Mail
Time	Telephone	Personal (in-home)	Mail
Cost	Mail	Telephone	Personal (in-home)
Accuracy			
Sampling control	Personal	Telephone	Mail
Supervisory control	Mail	Telephone (central location)	Personal
Opportunity for clarification	Personal	Telephone	Mail
Respondent Convenience	Mail	Telephone	Personal

Note: The comparative rankings of the personal, telephone, and mail survey techniques implied by this table are not universal; exceptions to these rankings can and do occur, depending on specific circumstances surrounding a research situation.

Exhibit 6.4 *Examples of Potential Research Constraints in Foreign Countries*

- Door-to-door interviews in Japan are conducted just inside or outside the home, rather than in a comfortable living room or den.
- In Japan each personal-interview participant expects to receive a gift, usually costing about $1.50 to $2.50.
- While virtually every household in the United States has a telephone, less than 50% of households in France have telephones.
- A questionnaire to be administered in Singapore must have an English version as well as equivalent Malay, Hokkien, Mandarin, Tamil, and Cantonese versions.
- Business executives and government officials in West Germany resist telephone interviews but are willing to be interviewed face to face.
- Several developed countries in Europe have no comprehensive lists of business establishments.
- The typical response rate for a mail survey conducted in India is 3% to 5%.

Source: The first two items are adapted from Andrew Watt, "A Day in the Working Life of a Japanese Interviewer Shows Some Similarities to U.S. and Some Differences," *Marketing News,* May 15, 1981. Section 2, p. 12; the next four items are adapted from David C. Pring, "Filling the Overseas Gaps," *Advertising Age,* October 26, 1981, p. 5; the last item is from a brochure entitled "MBA—Market Research," prepared by Marketing & Business Associates Pvt. Ltd., a leading marketing research firm in India.

ative summary of the three methods on the various criteria. As has been pointed out earlier, however, the rankings implied by this table will not necessarily hold in every research setting. But even if they did, no method can claim to be the best on *all* criteria. According to Robert Joselyn: "It is rare to encounter a research situation in which all considerations point to a single survey technique."[34] Or as Don Dillman correctly observes:

> Although each method has certain strengths and weaknesses, they do not apply equally, or sometimes not at all, to every survey situation. Thus, until the attributes of each method are considered in relation to the study topic, the population to be surveyed, and the precise survey objectives, the question of which is best cannot be answered.[35]

The purpose of our lengthy discussion concerning the relative merits and limitations of the three survey approaches was not so much to pinpoint the best approach; rather, it was to emphasize and illustrate the variety of factors to be considered in selecting the most suitable approach in a research project. Situation-specific considerations are particularly crucial in collecting primary data from countries other than the United States. Exhibit 6.4 offers a few interesting facts that illustrate unique research constraints that may have a significant bearing on the choice of an appropriate survey method and on the type of data that can be collected.

---

[34] Robert W. Joselyn, *Designing the Marketing Research Project* (New York: Mason/Charter, 1977).

[35] Dillman, *Mail and Telephone Surveys,* excerpted with permission from the publisher.

Nowadays, researchers are able to use hybrid or modified survey techniques that possess desirable features of the basic approaches and, at the same time, do not suffer seriously from their weaknesses. To illustrate, consider the shopping mall intercept interview, which, as we saw earlier, is a variation of the traditional in-home personal interview.[36] Shopping mall intercept interviews can be conducted at a faster rate than in-home interviews since interviewer travel is considerably reduced. They are also significantly less expensive.[37] Yet they can be conducted in such a way as to generate representative samples that are comparable to samples attainable through in-home or door-to-door interviewing.[38] Shopping mall interviews are also likely to be less intrusive and hence more convenient to respondents than in-home interviews.

Several commercial marketing research firms now have well equipped and sophisticated interviewing facilities in shopping malls. In such facilities, after potential respondents are intercepted and screened for eligibility by an interviewer, they can self-administer the survey questionnaire at a CRT terminal; as each question appears on the CRT screen, the respondent can directly type in his or her answer. To a certain degree, this technique offers respondents the privacy and flexibility of a mail questionnaire. It also offers researchers the speed, efficiency, and control of a computerized, central-location telephone survey. The example presented in Exhibit 6.5 illustrates the capabilities of sophisticated shopping mall research facilities.

TYPES OF OBSERVATION TECHNIQUES

We earlier looked at a few illustrations of using observation to collect data, before we discussed the various questionnaire techniques. The purpose of this section is to show the ways that observation studies can vary, depending on *how* the observations are made and *what* exactly is observed. The three observation methods presented in the Kwality Knitware case (methods D, E, and F) serve to illustrate some of these differ-

---

[36] For an overview of the use of this technique, see Roger Gates and Paul J. Solomon, "Research Using the Mall Intercept: State of the Art," *Journal of Advertising Research,* 22 (August/September 1982), pp. 43–49.

[37] See, for example, Howard N. Gundee, "MULTI-MALL Surveys Reduce Personal Interviewing Costs," *Marketing Today* (published by Elrick and Lavidge, Inc.), 21 (1) (1983); MULTI-MALL is a registered trademark of Elrick and Lavidge, Inc.

[38] Guidelines for carefully selecting shopping mall samples and adjusting for errors that may occur are discussed by Seymour Sudman, "Improving the Quality of Shopping Center Sampling," *Journal of Marketing Research,* 17 (November 1980), pp. 423–431; see also Alan J. Bush and Joseph F. Hair, Jr., "The Assessment of the Mall Intercept as a Data Collection Method," *Journal of Marketing Research,* 22 (May 1985), pp. 158–167.

Exhibit 6.5 *Sophisticated Shopping Mall Interviewing*

A major food manufacturer wishes to evaluate a new package design for one of its cereal products.

The research consultant has recommended the use of a shopping mall intercept technique where shoppers will be interviewed in the mall to find out their perceptions of and reactions to the new package design.

The marketing research firm has eight full-time shopping center facilities geographically dispersed across the United States. Each of the eight centers has a microprocessor unit with a video display and keyboard. To get a valid attitude measurement of the new package design, 400 interviews are needed. A total of 50 interviews will be completed in each of the eight facilities.

A questionnaire is designed in the headquarters office and programmed into the central computer. The eight shopping center facilities are then telephoned via a data-link line, and the questionnaire program is loaded into each microprocessor using a floppy disk. Different programs can be placed on a floppy disk and this allows program manipulation. The microprocessor is now ready for interviewing use.

Qualified consumers are intercepted in the mall, shown variations of the package design, and asked to self-administer the questionnaire. The individual actually sits at the terminal, reads the questionnaire as it appears on the video screen, and punches in responses. The data are automatically stored and tabulated in the microprocessor.

Once interviewing is completed, the central computer polls the results from the microprocessor at each shopping mall location, and data are read back to the central computer. Tabulated data are delivered to the client within three days, as compared to the two-week turnaround time that would probably have been required with less sophisticated methods.

Source: Reprinted by permission from *Business* magazine. "The New Technology for Market Researchers," by John T. Rougeau, July-September 1982, pp. 49–50.

ences. Recall that method D involves employing human observers, method E uses videotaping, and method F involves programming the cash registers to keep track of sales of Kwality brand shirts. The similarities and the differences across methods D, E, and F are discussed in the following subsections.

Natural Observation
A data collection approach in which events and/or behaviors are observed as they occur naturally.

Contrived Observation
A data collection approach in which events and/or behaviors are observed in an environment artificially set up by the researcher.

Natural Versus Contrived Observation

One feature common to methods D, E, and F is that they all involve making observations in a ***natural*** *setting*; that is, customer reactions and behavior are observed as they occur naturally in a real-life situation. Observations can also be made in a ***contrived*** *setting*, which is an environment artificially set up by the researcher.

An example involving contrived observation is the Litmus™ Simulated Test Marketing (LSTM) service offered by Yankelovich Clancy Shulman to predict how well a new product is likely to perform in the marketplace.

A key feature of the LSTM is a simulated, or contrived, store. The procedure used in LSTM research is as follows: A sample of potential customers first self-administers a background questionnaire dealing with demographics and buying habits. The respondents are then exposed to real advertising in a competitive context—for example, they may view an actual TV program containing commercials for the test product as well as for other products. Next, they are led in small groups to the store, where they are free to buy one or more of the products or none at all. Sales of the test product and other products in the same product category are observed. These sales data, along with additional feedback obtained from the respondents after they finish their shopping trip, are used to predict the market performance of the new product.[39]

An important advantage of contrived observation over natural observation is the greater degree of control the former offers. Such control permits researchers to collect relevant data in a speedy, efficient, and less expensive fashion. A potential drawback of contrived observation is the question of whether the collected data would result from a real-life setting. This limitation may be especially serious when respondents know they are being observed—which brings up another dimension on which observation methods can differ, namely, whether they are disguised or nondisguised.

Disguised Versus Nondisguised Observation

Disguised Observation
Observation in which respondents are unaware they are being observed.

Nondisguised Observation
Observation in which respondents are aware they are being observed.

An observation is ***disguised*** when respondents are unaware they are being observed; it is ***nondisguised*** when respondents are aware they are being observed. Monitoring sales at the cash register (method F) in the Kwality Knitware case is an illustration of disguised observation. Methods D and E can also be viewed as disguised observation if the researcher ensures that the human observers or videotaping equipment are not visible to the store's customers.

The main strength of disguised observation is that it allows for monitoring of the true reactions of individuals. Alternatively, data gathered through nondisguised observation may be contaminated by respondent-induced errors, just as in any questionnaire method. As Michael Ray points out: "Observation measures lose their advantage over interviews if the observation becomes apparent to respondents. This can happen with simple observation if the observer appears to be unusual, carries a clipboard, etc."[40] Since disguised observation is nonintrusive, it is also more convenient from the respondent's standpoint.

Serious ethical questions can arise, however, if disguised observation involves monitoring aspects of respondents' behaviors that normally are private or that may not be voluntarily revealed to researchers. Moreover, data gathered through a study using disguised observation may not be as rich as those obtained from a study involving nondisguised observation. Data from a nondisguised observation are richer because observation that is not hidden from respondents is generally done with their cooperation. Consequently, the respondents can be questioned before or after the observation to obtain additional data about characteristics that cannot be observed (e.g., data on demographics and buying habits collected in the LSTM procedure discussed earlier).

---

[39] Additional details about the LSTM are given in "Market Response Research and the Litmus System," a brochure prepared by Yankelovich Clancy Shulman, 8 Wright Street, Westport, CT 06880; also see Research in Use 8.1 in Chapter 8.

[40] Ray, *Unobtrusive Marketing*, p. 15.

Human Versus Mechanical Observation

Human Observation
A method of observation using human observers (as opposed to mechanical or electronic devices) to collect data.

Mechanical Observation
Observation in which mechanical or electronic devices (as opposed to human observers) are used to collect data.

Data collection in observation studies can be done by ***human** observers* or by ***mechanical** devices.* Method D in the Kwality Knitware example involves human observers; methods E and F use mechanical observation or, perhaps more appropriately, electronic observation. With the advent of increasingly sophisticated technological gadgets, mechanical observation is rapidly replacing human observation. Of course, irrespective of how observations are made, *interpretation* of those observations is very much a human responsibility.

A key strength of mechanical observation is that it can be used to monitor, in a more precise fashion, virtually anything a human observer is capable of monitoring. For instance, videotaping customers in the special display area for Kwality brand shirts (method E) is likely to generate more complete and accurate observations than using human observers (method D). Moreover, method E is also likely to be less conspicuous to customers than method D—an important consideration if disguised observation is crucial.

Numerous mechanical and electronic devices are now available to observe things human observers may find very difficult, if not impossible, to study. A comprehensive discussion of such devices is beyond the scope of this text. However, several illustrations are provided next.

Eye-Tracking Equipment. This equipment is used to ascertain precisely which sections of an ad, product packaging, or promotional display attract customers' attention, and how much time they spend looking at those sections.[41]

Response Latency
The speed with which a respondent provides an answer.

Devices to Measure Response Latency. ***Response latency*** is the speed with which a respondent provides an answer. Response latency measures are being increasingly employed to determine the effectiveness of ads in influencing the strength of consumers' brand preferences. The rationale underlying the use of response latency measures is that the

---

[41] For further details about the capabilities of eye-tracking equipment, see, for example, Bernie Whalen, "Eye Tracking Technology to Replace Day-After-Recall by '84," *Marketing News*, November 27, 1981, Section 1, p. 18; Elliot Young, "Use Eye Tracking Technology to Create Clutter-Breaking Ads," *Marketing News*, November 27, 1981, Section 1, p. 19.

quicker a respondent expresses preference for a brand, the *stronger* the preference is. For instance, when two customers are asked whether they prefer 7-Up or Sprite, they may both say 7-Up; however, the one that responds faster, even by a fraction of a second, presumably has a lower degree of doubt in his or her mind and hence a stronger preference for 7-Up. Only electronic devices can accurately monitor and detect differences in response latency.[42]

Voice Pitch Analysis (VOPAN) *A technique for assessing how much emotional commitment is attached to a respondent's verbal answers; it involves measuring the voice pitch of verbal responses during an interview and comparing it with the respondent's normal voice pitch during routine conversation about neutral topics.*

Instruments to Conduct Voice Pitch Analysis (VOPAN). ***Voice pitch analysis,*** not unlike response latency measures, is used to determine how strongly a respondent feels about an answer or how much emotional commitment is attached to it. The VOPAN involves measuring the voice pitch of verbal responses during an interview. This voice pitch is then compared with the respondent's normal voice pitch used during routine conversation about neutral topics such as the weather. The extent to which the voice pitch of responses to a survey question deviates from normal voice pitch is considered a measure of the respondents' emotional commitment to the answer.[43]

Audimeter *An electronic device used by the A. C. Nielsen Company to monitor television-viewing behavior—what channels are watched, when, and for how long.*

Audimeter. As we saw in Chapter 5, the ***audimeter*** is an electronic device used by the A. C. Nielsen Company to monitor television-viewing behavior—what channels are watched, when they are watched, and for how long they are watched. The audimeter is attached to the television sets of a representative panel of households across the United States. The data collected by the audimeters are automatically transferred to a central computer for analysis and interpretation.

People Meter *An electronic device used by the A. C. Nielsen Company to monitor television-viewing behavior—who is watching, in addition to what shows are being watched.*

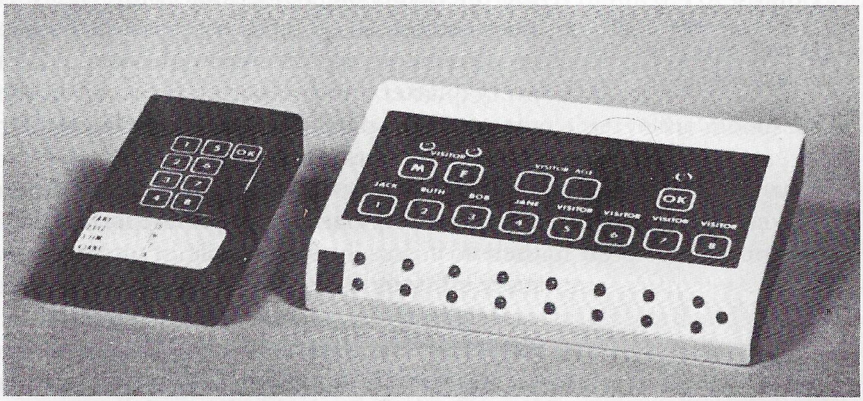
People Meter. The latest development in the area of monitoring television-viewing behavior is the introduction of the ***people meter***—an electronic device designed to monitor *who* is watching TV in addition to what show is being watched. The people meter uses the latest electronic technology and has separate buttons for all household members as well as for visitors. The individual viewers simply press their corresponding buttons on the people meter when they start or stop watching TV. A picture of a people meter used by the A. C. Nielsen Company is shown in Exhibit 6.6. All the data collected by Neilsen's people meters are automatically transmitted to its central computer.

---

[42] Several recent articles discuss response latency measures and their use in marketing research; see James MacLachlan and John G. Myers, "Using Response Latency to Identify Commercials That Motivate," *Journal of Advertising Research,* 23 (October/November 1983), pp. 51–57; David A. Aaker, Richard P. Bagozzi, James M. Carman, and James M. MacLachlan, "On Using Response Latency to Measure Preference," *Journal of Marketing Research,* 17 (May 1980), pp. 237–244.

[43] For further discussions of VOPAN and its use, see, for example, Ronald Nelson and David Schwartz, "Voice Pitch Gives Marketer Access to Consumer's Unaware Body Responses," *Marketing News,* January 28, 1977, p. 21; Nancy J. Nighswonger and Claude R. Martin, Jr., "On Using Voice Analysis in Marketing Research," *Journal of Marketing Research,* 18 (August 1981), pp. 350–355.

Exhibit 6.6 *Nielsen's People Meter*



Source: A. C. Nielsen Company, Nielsen Plaza, Northbrook, Ill. 60062. Courtesy of A. C. Nielsen Company.

While the people meter has the potential for providing highly refined television-audience measurements, there is some controversy surrounding its use. Specifically, for some TV programs, significant discrepancies in the estimates of audience size and composition have surfaced between those provided by people meters and those provided by the traditional diary method, wherein households participating in a TV panel make entries into a diary regarding who watched each program. Consequently, questions have been raised about the validity of ratings based on people meter data and the representativeness of households agreeing to have their television-viewing behavior monitored through people meters. At the present time efforts are being made to resolve these questions and further refine the people meter methodology.[44]

Direct Versus Indirect Observation

Suppose Mr. Smith, owner of a gift shop, wants to know how much customer traffic his store attracts during the last quarter of a year relative to the first three quarters. One option available to Mr. Smith is to use human observers or mechanical/electronic monitors to count the number of people entering his store during each quarter. This procedure is ***direct observation*** since the actual behavior or phenomenon of interest

Direct Observation *Observation that involves looking at a phenomenon rather than the phenomenon's consequences.*

---

[44] For more information on the controversy surrounding people meters, see the following: Peter Barnes and Joanne Lipman, "Networks and Ad Agencies Battle over Estimates of TV Viewership," *Wall Street Journal,* January 7, 1987, p. 23; David Lieberman, "The Networks' Big Headache—Nielsen's New 'People Meter' Could Batter Their Ad Revenues," *Business Week,* July 6, 1987, pp. 26–28; David F. Poltrack, "Living with People Meters," *Journal of Advertising Research,* June/July 1988, pp. RC-8–RC-10.

***Indirect Observation** Observation that involves looking at the results or consequences of a phenomenon rather than directly observing the phenomenon itself.*

is observed. Another option for Mr. Smith is to use some form of ***indirect observation*** which consists of examining the results or consequences of the phenomenon. For instance, Mr. Smith could install, just inside the store entrance, a temporary patch of floor covering specially designed to wear out easily and show the extent of wear. Replacing the patch every quarter and comparing the wear on the fourth-quarter patch with the wear on patches used in previous quarters can help answer Mr. Smith's question.

Indirect observation, when compared with direct observation, can give only relatively crude or imprecise indications of a phenomenon. Nevertheless, it does have a couple of attractive features. It will generally be more efficient than direct observation from a time and a cost perspective, especially if the behavior to be observed is stretched out over a long period of time and is likely to occur infrequently during that time. As we saw earlier, monitoring sales of Kwality brand shirts at the cash register (method F)—an indirect-observation method—will be less time-consuming and expensive than methods D and E, both of which involve direct observation. Other examples of efficient indirect observation include examining the wear and tear on magazines to estimate their readership and analyzing the accumulation of fingerprints on advertisements to ascertain the amount of consumer exposure to them.[45]

Another advantage of indirect observation is that it may be the only way of getting relevant data from situations that are impractical to observe directly. To illustrate, Information Data Search, Inc., (IDS), a competitive intelligence firm, uses a number of creative indirect-observation techniques, one of which is cardboard box analysis. Leonard M. Fuld, managing director of IDS, describes this technique as follows:

> You can learn a great deal about a company just by looking at the boxes it uses to ship its products to retailers. The bottom of the box has a seal, and numbers, and the name of the cardboard manufacturer. . . . So all you do is call up the suppliers and talk to them. Usually they'll tell you how many they produced for a company in a certain period. That information can be used to generate pretty reliable estimates of production runs, unit shipments, sales—you name it.[46]

***Structured Observation** Observation in which the data gathered pertain to discrete, clearly defined events or behaviors and are therefore recorded by merely making entries or checking off categories on an appropriate observation form.*

Structured Versus Nonstructured Observation

Whether an observation is structured or nonstructured depends on the nature and type of data recorded. Observation is *structured observation* when a study's data requirements are well-established and can be broken into a set of discrete, clearly defined categories—much like the precoded response categories in structured questions.

Consider a restaurant manager who wants to know the number of single customers and the number of parties of two or more customers who eat at the restaurant. This setting is eminently suited for structured observation since the data requirements are clearly defined. For instance, waiters and waitresses can be used as observers and asked to keep a tally of the number of single customers and the number of groups of customers they serve.

---

[45] For an excellent discussion of these and other indirect-observation techniques, see Webb, Campbell, Schwartz, and Sechrest, *Unobstrusive Measures,* Chapter 2.

[46] Bernie Whalen, "Marketing 'Detective' Reveals Competitive-Intelligence Secrets," *Marketing News,* September 16, 1983, Section 1, p. 1.

Nonstructured Observation *Observation in which everything of relevance is observed and recorded in an open-ended fashion.*

Suppose, however, that the manager wants to observe the moods and behaviors of single customers and of customer groups. Given the vagueness of the manager's objective, some form of ***nonstructured observation*** will be required in this situation. For instance, the manager may use inconspicuous videotaping equipment or human observers to monitor "everything of relevance" concerning the restaurant's customers.

While structured observation is generally easier to record and analyze than nonstructured observation, it is limited in terms of the depth and richness of data it can provide. These pros and cons of structured observation are similar to those of structured questionnaires discussed earlier. Along the same lines, structured observation is more suitable for conclusive-research projects than for exploratory-research projects.

A final note on the observation method: What we have discussed so far under "Types of Observation Techniques" are really five dimensions along which observation methods can vary. Numerous types of observation methods are possible from unique combinations of these dimensions. Differences between observation methods, and the extent of those differences, depend on how the methods vary along the five dimensions. For example, method D in the Kwality Knitware case can be labeled as natural, disguised (if the observer is not conspicuous), human, direct, nonstructured observation. In contrast, method F is natural, disguised, mechanical, indirect, structured observation.

SUMMARY

This chapter described a number of primary-data collection techniques that can be broadly classified into questioning and observation methods. In general, the questioning method is more versatile than the observation method, mainly because the latter can only be used to study what is visible. The questioning method is also typically, but not always, faster and cheaper. However, the observation method, since it is less likely to suffer from respondent-induced errors, is capable of generating more-accurate data. The observation method is also more convenient from the respondent's perspective. In some studies one may be able to use a combination of the two methods in such a way as to capitalize on the advantages of both.

The format of questionnaires can vary along two dimensions, structure and disguise. The more conclusive, or final, a research project, the more appropriate and advantageous structured questionnaires are. The need to use disguised questionnaires increases with anticipated respondent unwillingness and/or inability to answer direct questions.

open-ended questions, for example

On the basis of structure and disguise, questionnaires can be roughly classified into four broad categories. First are structured, nondisguised questionnaires, which are frequently used, especially in descriptive-research projects. Second are nonstructured, nondisguised questionnaires, which are particularly appropriate for exploratory-research projects. Third are nonstructured, disguised questionnaires, which usually employ projective techniques and are used in motivation research studies to explore consumers' inner feelings and motives. Fourth are structured, disguised questionnaires, whose typical purpose is to uncover people's attitudes toward broad social issues; because of their purpose, they are not frequently used in traditional marketing research studies.

) word association test
2) sentence completion test
3) TAT picture
→ story
4) cartoon test

Questionnaires can also be classified into three categories according to the method used to administer them: personal interviews, telephone surveys, and mail surveys. These questionnaire administration methods differ in terms of their versatility, speed, cost, accuracy, and convenience from the respondent's standpoint. That any one method will be the best on *all* criteria in a given situation is highly unlikely. Therefore the researcher must carefully consider situation-specific characteristics and ascertain which criteria are most critical before selecting a questionnaire administration method. At times, the researcher may wish to improvise a hybrid questionnaire administration method possessing key strengths of all the basic methods.

A variety of techniques are available for collecting data through observation. Specifically, an observation study can be (1) conducted in a natural or a contrived setting; (2) disguised or nondisguised from respondents; (3) conducted by human observers or by mechanical/electronic means (the latter are becoming increasingly popular because of technological advances); (4) direct or indirect, depending on whether an actual occurrence or its consequences are observed; and (5) structured or nonstructured. As in the selection of a questionnaire method, the choice of an appropriate observation technique must be based on a careful examination of factors such as the nature of the research, the type of data required, and the resources available.

(look for more detail)

KEY TERMS

Primary data
Questioning
Observation
Questionnaire
Completely structured question
Completely nonstructured question
Open-ended question
Disguised question

extremes

Nondisguised question
Depth (in-depth) interview (part of nons, nond)
Motivation research
Projective technique
Word association test
Sentence completion test
Thematic apperception test
Cartoon (balloon) test

Personal interview
Telephone survey
Mail survey
Shopping mall intercept interview
Computer-assisted telephone interviewing (CATI)
Natural observation (real)
Contrived observation (fixed situation)

Disguised observation
Nondisguised observation
Human observation
Mechanical observation

(time, speed)
Response latency
Voice pitch analysis (VOPAN)
Audimeter (what channels)
People meter (who watches)

Direct observation
Indirect observation
Structured observation
Nonstructured observation

QUESTIONS

1. Can observation methods be used in exploratory research? Why or why not?
2. In what ways is the questionnaire method more versatile than the observation method?
3. The manager of a large independent supermarket in a city wants to estimate the proportion of households in the city that make a purchase in the supermarket at least once a month. What method(s) should the manager use to generate this estimate? Explain your answer.
4. A suburban shopping mall is planning to exhibit antiques and various types of art in its lobby area to generate more customers for the mall. Describe a *combination* of the questionnaire and observation methods that can be used to ascertain the effectiveness of the antique and art exhibit.
5. What are *depth interviews?* In what situations are they likely to be most useful? Explain your answer.
6. Describe the distinct features that all projective techniques have. Pick any one of the several projective techniques discussed in the chapter, and give an example of your own to illustrate a real-life situation in which the technique will be useful.
7. Discuss the accuracy of the following statement: "Nondisguised questionnaires are much more widely used in marketing research today than disguised questionnaires."
8. What advantages do telephone surveys have over personal interviews and mail surveys?
9. A state university wants to do a study of graduate students' attitudes about its summer school calendar. What survey method would you recommend? Why?
10. What is a *shopping mall intercept interview?* What are its advantages over the more traditional questionnaire administration methods?
11. Recall the Trent Eating Association (TEA) situation presented in Chapter 4 (situation C at the beginning of Chapter 4). Assume that an experimental study is to be conducted in this situation and that the necessary data are to be gathered through observation. Describe what this observation method would look like along each of the five observation dimensions discussed in this chapter.

CHAPTER 7

Qualitative Research

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between qualitative and quantitative research and explain the relationship between the two.*
- *State and illustrate the three major determinants of focus group effectiveness.*
- *Name the key skills that an effective focus group moderator should possess.*
- *Describe the major advantages and drawbacks of focus groups.*
- *Discuss, with suitable examples, potential applications of focus groups.*
- *Describe the distinct features of dual-moderator groups, dueling-moderator groups, crowded one-on-one interviews, and shopper studies.*

WHAT IS QUALITATIVE RESEARCH?

Qualitative Research
A form of exploratory research involving small samples and nonstructured data collection procedures.

Qualitative research involves collecting, analyzing, and interpreting data that cannot be meaningfully quantified, that is, summarized in the form of numbers. For this reason qualitative research is sometimes referred to as *soft* research. This term is unfortunate, because as subsequent sections of this chapter will show, soft research is no less valuable than so-called hard, or quantitative, research.

Any study using *nonstructured*-questioning or -observation techniques can be labeled as qualitative research. However, qualitative research typically involves relatively few respondents or units. In other words, a study involving a large representative sample would normally not be called qualitative research even if it used some nonstructured questions or observations.

The nonstructured and small-sample features of qualitative-research techniques have an important implication concerning their application: Such techniques are intended to provide initial insights, ideas, or understanding about a problem; they are *not* meant to recommend a final course of action. Therefore qualitative-research techniques are most appropriate in situations calling for exploratory research. Several techniques covered in previous chapters—for example, the key-informant technique, the case study method, the focus group interview, the in-depth interview, the word association test, and nonstructured observation—are qualitative-research techniques that are often used in exploratory research.

QUALITATIVE VERSUS QUANTITATIVE RESEARCH

Quantitative Research
A form of conclusive research involving large representative samples and fairly structured data collection procedures.

Quantitative research, in contrast to qualitative research, is characterized by more structure and by larger, more representative respondent samples. Consequently, the logical place for quantitative-research techniques (usually in the form of large-scale questionnaire surveys or structured observations) is in conclusive-research projects.

Table 7.1 outlines several illustrative situations to highlight the distinction between qualitative- and quantitative-research applications. A common thread running through the research settings under "Qualitative Research" in Table 7.1 is the need for developing an initial understanding of something. A primary role of qualitative research is to generate hunches or hypotheses that can be tested through more formal research. In contrast, each of the situations under "Quantitative Research" calls for very specific data, capable of suggesting a final course of action. A primary role of quantitative research is to test hunches or hypotheses. Thus differences in methodology (e.g., sample selection, data collection, data analysis) between qualitative- and quantitative-research studies parallel methodological differences between exploratory and conclusive research discussed in Chapter 4 (see Table 4.1).

IMPORTANCE OF QUALITATIVE RESEARCH

We saw in Chapter 1 that there is a lot more to marketing research than numbers and statistics. This point needs to be reemphasized, since it is tempting to believe (and some *do* believe) that qualitative research is somehow inferior to quantitative research and hence is truly not marketing research. However, quantification by itself does not make research any more accurate or valuable than soft research. One marketing research practitioner had the following reaction to those who refuse to recognize the importance of anything but quantitative research:

> The line between qualitative and quantitative research is not nearly as distinct as the numbers people would have us believe, consistent with a discipline which is more art than science. Both types of research have drawbacks which we must account for in deriving maximum value from their use. . . . Quantitative research is subjective, sometimes highly so, though admittedly generally less so than qualitative. If this is true, where then do we draw the line between marketing research and "something else"? . . . Numbers don't speak; analysts do. Figures lie and liars figure.[1]

---

[1] "Qualitative *Is* Marketing Research Because It Aids Decision Makers, Helps Reduce Risk," *Marketing News,* January 22, 1982, Section 2, p. 8.

Table 7.1 *Applications of Qualitative and Quantitative Research*

	Examples of Research Settings Best for	
Topic Area	**Qualitative Research**	**Quantitative Research**
Advertising	The marketer of Shine detergent wants to come up with ideas for creatively communicating the detergent's benefits through a television commercial	Two different commercials for Shine detergent have been developed; management wants to know which of the two will be more effective in favorably influencing target customers' preferences for Shine
Product planning and promotion	The brand manager in charge of Rise baking soda wishes to develop an understanding of how, when, where, and why consumers use Rise	The brand manager wants to know whether distributing a 20-cents-off coupon for Rise will significantly increase its unit sales
Personal selling	The sales of a firm employing a large sales force have been dropping steadily; the sales manager wants to identify possible reasons for this decline	From discussions with several salespeople the sales manager suspects that the sales force's morale is low and wants to confirm it by using a standard employee morale questionnaire
Services marketing	The administrators of a hospital want to develop a feel for the nature and the extent of apprehension experienced by patients when they are in the hospital	The administrators want to ascertain patients' ratings of specific attributes of such things as room service, food service, and medical treatment
Politics	A candidate running for a U.S. Senate seat from Alaska wants to understand the social, economic, and environmental concerns of Alaskans	The candidate wants to determine the percentage of registered Alaskans that favor him over opponents vying for the same Senate seat

Similar sentiments have been expressed by other practitioners. For instance, according to a director of marketing research for the Pillsbury Company:

> While these two terms [qualitative and quantitative] distinguish between two different forms of research, the dichotomy isn't quite as clearcut as some would think . . . All data are "soft" and any way of gathering data is a series of compromises. What's key in commercial research is transforming data into information, insights, and knowledge.[2]

And in the words of Ernest Dichter, a pioneer in the area of applying motivation research techniques to marketing: "Too much [quantitative] attitude research is oversimplified. It counts noses instead of trying to explore the complexity of the human soul."[3] Some researchers have even suggested that qualitative research is needed to verify the accuracy of quantitative research results and to interpret them properly.[4] According to Rohit Deshpande: "Frequently, statistical results from survey data analysis can be validated by recourse to qualitative observations and informant interviews."[5]

In fairness to the critics of qualitative research, however, we must mention that qualitative research has been misused by careless, and at times unscrupulous, researchers.[6] The most frequent misuse is making generalizations from research that is intended primarily to provide preliminary insights about a given research setting.

The proper position to take in a qualitative-versus-quantitative debate is to view *both* types of research as playing a legitimate and important role in marketing research. Just as exploratory research is a necessary prelude to conclusive research in most situations, qualitative research can be viewed as generating just the right type of raw material needed to produce a finished product in the form of a relevant quantitative-research project. Carrying this product analogy further, to view the raw material as the finished product is inappropriate and ridiculous. Similarly, without suitable raw material the finished product is likely to be unsatisfactory and even dangerous. The point is that in a properly conceived study qualitative- and quantitative-research techniques must *complement,* rather than *compete with,* each other.

---

[2] Roger E. Bengston, "Despite Controversy, Focus Groups Are Used to Examine Wide Range of Marketing Questions," *Marketing News,* September 19, 1980, p. 18.

[3] Ernest Dichter, "Try These 10 Research Techniques to Uncover *Real* Attitudes," *Marketing News,* May 13, 1983, Section 2, p. 6.

[4] F. D. Reynolds and D. K. Johnson, "Validity of Focus Group Findings," *Journal of Advertising Research* (June 1978), pp. 21–24; see also Johny K. Johansson and Ikujiro Nonaka, "Market Research the Japanese Way," *Harvard Business Review,* May-June 1987, pp. 4–7, for an insightful discussion of how Japanese executives place much more emphasis on qualitative than on quantitative research.

[5] Rohit Deshpande, " 'Paradigms Lost': On Theory and Method in Research in Marketing," *Journal of Marketing,* 47 (Fall 1983), p. 108.

[6] See, for example, "Qualitative Research Isn't Marketing Research; New Name May Promote Proper Use of Tools," *Marketing News,* January 22, 1982, Section 2, p. 9.

We now turn to focus group interviews, which are perhaps the most widely and frequently used form of qualitative research.

FOCUS GROUP INTERVIEWS

As stated earlier, we have already covered several qualitative-research techniques. One, mentioned in Chapter 4, is the ***focus group interview,*** often simply called focus group. This technique is such a popular type of qualitative research that many marketing research practitioners consider it synonymous with qualitative research. (As evidence of this practice, most articles on the topic appearing in *Marketing News* use the two terms interchangeably.)

A survey of business firms conducted in the midseventies revealed that about 50% used focus groups in their marketing research.[7] Also, a 1987 study of a sample of the largest consumer product companies in the United States showed that 98% used focus groups, up from 90% in a similar study conducted in 1983.[8] Finally, according to a 1981 survey of commercial marketing research firms in the United States, approximately 60% of them frequently conducted focus group interviews for their clients.[9] Clearly, there is a prevalence of focus groups in marketing research practice.

A significant portion of the remainder of this chapter is devoted to focus groups—not only because of their popularity but also because certain methodological issues related to focus groups are relevant to other qualitative-research techniques as well.

Definition

A good way to get a feel for what goes on in focus groups interviews is to examine the actual procedure followed by a firm that conducts a large number of them. The following description provided by Burke Marketing Research, Inc., serves as a useful introduction to focus group interviews.

> The respondents are requested to come to the Burke studios at a specified time. There the discussion is conducted in a relaxed atmosphere, with the environment approximating that found in a home, rather than those of

---

[7] Barnett A. Greenberg, Jac L. Goldstucker, and Danny N. Bellenger, "What Techniques Are Used by Marketing Researchers in Business?" *Journal of Marketing,* 41 (April 1977), pp. 62–68.

[8] "Practices, Trends and Expectations for the Market Research Industry," a report prepared (April 29, 1987) by Market Facts, Inc., 676 North St. Clair Street, Chicago, Ill. 60611.

[9] A. Parasuraman, "A Study of Techniques Used and Clients Served by Marketing Research Firms," *European Research,* 10 (October 1982), pp. 177–185.

> an office or institution. Placing the respondents in a normal, relaxed situation is an important factor in gaining the prompt cooperation of the individuals in the group as the discussion begins. Interviewing is done during the day, if only non-working housewives are required, or at night if working women or men are needed.
>
> Burke discussions usually begin with the session leader explaining to the respondents the general purpose of group interviewing and the importance of their candid opinions. The discussion leader will normally have a written outline, which has been prepared to make sure that all subjects relevant to the objectives of the research will be covered. This outline proceeds from the general to the specific. Very often the leader, having brought up the broad general subject, will need to do very little guiding to see that all subject areas are covered. Usually these will be covered in any order brought up by respondents rather than following the written outline. This promotes normal conversation and the discussion leader only needs to probe into details within each area and to keep bringing out those respondents who are reluctant to voice their feelings. The discussion leader's own attitudes, of course, must never influence the discussion.[10]

This procedure is not unique to focus group interviews conducted by Burke; rather, it is a fair description of focus groups in general. It also suggests the following concise definition.

Definition. A ***focus group interview*** involves an objective discussion leader or moderator who introduces a topic to a group of respondents and directs their discussion of it in a nonstructured and natural fashion.

The air of informality implied in this definition does not mean that focus group interviews are easy to conduct. Quite to the contrary, focus groups as well as other qualitative-research techniques must be conceived and conducted with the utmost care if they are to be useful. As Bobby Calder correctly observes, "Qualitative marketing research is more complex than any simple notion that quantitative research permits objective numerical analysis which qualitative research sacrifices for intensive analysis and fast turnaround."[11]

The potential worth of focus group interviews depends highly on factors such as group composition, moderator characteristics, and the interview atmosphere itself. Unfortunately, there is no unanimous agreement among researchers about how one should select or manipulate these factors in order to reap maximum benefits from focus group interviews. According to the authors of a comprehensive monograph on qualitative research: "The interviews conducted with researchers and the readings examined in the preparation of this monograph revealed a number of disagreements concerning the use and methods involved in conducting research using focus group interviews."[12] Therefore remember that any general inferences about focus groups made in the following sections, while reflecting the opinions of a majority of researchers, are not absolute.

---

[10] "Group Interviewing," an internal manual (updated) prepared by Burke Marketing Research, Inc., 800 Broadway, Cincinnati, Ohio 45202, p. 7.

[11] Bobby J. Calder, "Focus Groups and the Nature of Qualitative Marketing Research," *Journal of Marketing Research,* (August 1977), p. 360.

Group Composition

The effectiveness of a focus group depends very much on the number of participants and their characteristics. There is no scientific basis for determining the optimum focus group size. However, virtually all focus groups conducted by marketing research practitioners involve between 8 and 12 participants. Since focus groups have been in use for many years, and since practitioners are likely to continue to follow a guideline only if they find it useful, we can infer that the 8-to-12 group size is a rule of thumb. Groups with less than 8 participants are apparently not likely to generate the momentum and group dynamics necessary for a truly beneficial focus group session (benefits unique to well-conducted focus groups will be discussed later). Similarly, a group with more than 12 participants may be too crowded to be conducive to a cohesive and natural discussion.[13]

Regarding the characteristics of participants, there is considerable agreement among published articles discussing this issue that a focus group must be as *homogeneous* as possible with respect to demographic and socioeconomic characteristics. "It is essential to get as much commonality in a group as possible so that the numerous interacting demographic variables do not confuse the issues; to be most productive, all the participants must be on the same wave length."[14]

Conducting focus groups using similar respondents has important implications concerning the generalizability and truthfulness of the collected data. An example will help clarify these implications.

---

[12] Danny N. Bellenger, Kenneth L. Bernhardt, and Jac L. Goldstucker, *Qualitative Research in Marketing* (Chicago: American Marketing Association, 1976), p. 7. This monograph provides an excellent discussion of qualitative-research techniques; several of the ideas developed in this section stem from it. Other books that offer guidelines for conducting focus group interviews include Alfred E. Goldman and Susan Schwartz McDonald, *The Group Depth Interview* (Englewood Cliffs, N.J.: Prentice-Hall, 1987); Thomas L. Greenbaum, *The Practical Handbook and Guide to Focus Group Research* (Lexington, Ma.: D.C. Heath, 1988); and Jane Farley Templeton, *Focus Groups: A Guide for Marketing and Advertising Professionals* (Chicago: Probus, 1987).

[13] That rules of thumb are just that and not universal laws is emphasized in at least one controlled study (as opposed to studies involving focus groups conducted in practice), which found that two four-member groups are better than one eight-member group; see Edward F. Fern, "The Use of Focus Groups for Idea Generation: The Effects of Group Size, Acquaintanceship, and Moderator on Response Quantity and Quality," *Journal of Marketing Research,* 19 (February 1982), pp. 1–13.

[14] Bellenger, Bernhardt, and Goldstucker, *Qualitative Research in Marketing,* p. 9.

Example. Beauty Products, Inc., (BPI) is a firm marketing a variety of women's cosmetics. BPI is currently seeking ideas or concepts for developing a new cosmetic line with products appealing to a wide cross section of women. One logical starting point for BPI is to conduct focus group interviews with women to get them to talk freely about their makeup needs, use of cosmetics, satisfaction and dissatisfaction with existing products, suggestions for improvements, and so on. BPI decides to conduct a focus group interview of working women between the ages of 20 and 30.

Clearly, whatever inferences BPI draws from the focus group information cannot be generalized. There is no guarantee that ideas for new cosmetic products stemming from this group will appeal to other groups of women. This limitation is present in virtually all focus groups since they typically involve small homogeneous samples. This limitation does not mean that focus groups are useless, though. The purpose of a focus group—as a matter of fact, the purpose of *any* qualitative-research technique—is to generate preliminary insights rather than to make generalizations.

The researchers for BPI would do well to conduct a *series* of focus groups, each internally homogeneous but different from the others in terms of age and employment status of the participants. Such a series of focus group interviews is likely to generate a range of new-product ideas appealing to a wide cross section of women. Indeed, judging from what is done in practice, very seldom is just one focus group conducted for a research project. A rule of thumb is to conduct at least two focus groups aimed at a single research issue to generate a meaningful, interpretable set of insights. Ideally, resources permitting, additional focus groups should be conducted as long as new information continues to emerge from them.

Though conducting several focus groups may yield a rich reservoir of data, the fact still remains that it is dangerous to draw any definitive conclusions from such data. This point is important because making generalizations on the basis of focus group interviews is perhaps the most frequent misuse of the technique. As one practitioner recently observed:

> [An] abuse [related to focus groups] is accumulating a large sample by a series of groups and trying to quantify the data. The need for qualitative or quantitative data should be determined before commitment to the technique. . . . Information from 100 respondents in 10 focus groups still isn't representative, because of the strong sampling and group dynamics biases. Information from 100 respondents interviewed individually might well be representative.[15]

---

[15] Martin M. Buncher, "Focus Groups Seem Easy to Do and Use, but They're Easier to Misuse and Abuse," *Marketing News*, September 17, 1982, Section 2, p. 14.

Let us now examine the issue of the extent to which responses generated in a focus group are truthful or trustworthy. As we saw in the previous chapter, data obtained through questioning respondents can be biased owing to a variety of factors, such as respondent untruthfulness and interviewer errors. Focus group interviews are no exception to this bias, especially since participants in a group interview may be tempted to provide biased responses to impress others, to appear socially responsible, and so on. Moreover, a group that does not look "natural" from the participants' perspective may lead to forced or contrived responses, thereby aggravating the problem of response inaccuracies.

Intuitively, a homogeneous group in which participants can identify and feel comfortable with one another is likely to be more natural and relaxed than a heterogeneous group. Therefore contrived responses or responses intended merely to impress others are less likely to arise in a homogeneous group. To illustrate this point, let us once again look at the BPI case.

What would you expect to happen if BPI used heterogeneous rather than homogeneous focus groups in its quest for new-product ideas? Specifically, suppose a focus group interview is to be conducted with working and nonworking women, drawn from a wide range of age-groups. Since working status and age can strongly influence perceptions about the use of cosmetics, the participants in this group will most likely hold widely varying, and perhaps conflicting, opinions. Furthermore, differences in working status and age may lead to undesirable consequences; for instance, some women may become reticent, some may provide socially acceptable responses, and others may attempt to show off. The result is likely to be an unproductive focus group session, in terms of both the number of relevant insights generated and the validity of those insights. Thus homogeneous focus groups, although not capable of guaranteeing completely valid responses, are less likely to suffer from response inaccuracies than heterogeneous groups are.

The preceding discussion emphasizes the need to use proper screening procedures to ensure suitable group composition. Demographic similarity among participants in a group is desirable from the standpoint of generating useful, trustworthy data. An easy way of accomplishing this similarity is to recruit respondents from organizations such as church groups, clubs, and professional associations whose members are likely to be similar. A potential problem here, however, is whether this selection procedure will lead to any detrimental group behavior. According to Bellenger, Bernhardt, and Goldstucker: "Many researchers believe that an individual should not be allowed to participate in a group containing a friend, neighbor, or relative; they will tend to talk to each other and not to the group as a whole."[16]

---

[16] Bellenger, Bernhardt, and Goldstucker, *Qualitative Research in Marketing*, p. 10.

However, controlled studies have not demonstrated any definite problem associated with having acquaintances and friends in the same group.[17] Moreover, several well-known commercial marketing research firms apparently do recruit focus group participants from organizations like churches and schools.[18] Such recruiting can be justified, since a well-qualified moderator can usually overcome difficulties such as participants talking to each other and not to the group as a whole.

Professional Respondent
A person who has already participated in numerous focus group interviews.

A problem more serious than participants who know one another beforehand is the problem of so-called ***professional respondents***, people who have participated in numerous focus groups before. Professional respondents are atypical consumers, and their responses during a group session can lead to serious validity problems.[19] While one is screening respondents for focus groups, a good practice is to exclude those who have participated in a focus group within the previous six months or so.

Last but not least, recruited participants must have had some experience related to the product or the issue to be discussed.[20] The reason should be obvious: Respondents lacking relevant experience are not likely to make any valuable contribution to the group discussion. To the contrary, they may adversely affect the discussion by making meaningless remarks just for the sake of participating. Therefore the screening procedure should query potential respondents about their past experience with or exposure to the focus group topic.

Exhibit 7.1 shows a screening questionnaire used by a research firm to recruit female respondents via telephone for a focus group dealing with appliance repair services. The first question is intended to ensure that potential respondents have had relevant experience. The second and third questions reflect client-imposed restrictions regarding participant characteristics. The last two questions are meant to screen out atypical and professional respondents. Only individuals who pass all five questions are recruited to participate in the focus group.

Moderator Characteristics

Of all the factors influencing the effectiveness and usefulness of a focus group, the discussion leader or moderator is perhaps the most crucial. The moderator's role, in addition to being crucial, is an extremely delicate one. It involves stimulating natural discussion among all the participants while at the same time ensuring that the focus of the discussion does not stray too far from the topic.

---

[17] See Fern, "The Use of Focus Groups for Idea Generation."

[18] Burke Marketing Research, Inc., is one such group; see "Group Interviewing," p. 5.

[19] For an interesting and insightful discussion of problems stemming from professional respondents, see Hazel Kahan, " 'Professional' Respondents Say They're Better for Research than 'Virgins,' but They're Not," *Marketing News,* May 14, 1982, Section 1, p. 22.

[20] Bellenger, Bernhard, and Goldstucker, *Qualitative Research in Marketing,* p. 10.

Exhibit 7.1 *Screening Questionnaire to Recruit Focus Group Participants*

NAME:________________DATE:________________

ADDRESS:________________PHONE:________________

CITY & STATE:________________INTERVIEWER:________________

Hello, this is __________ from Chicagoland Field Service. We are planning a group discussion on the topic of appliance repair and service. Would you be interested in participating in such a session on __________ at __________ ? (IF "NO" terminate and tally)

1. Have you had an appliance (such as a television, washer, dryer, toaster, etc.) serviced or repaired within the last year?

YES .... 1 CONTINUE

NO .... 2 TERMINATE AND TALLY

2. Are you between the ages of 30 and 64?

YES .... 1 CONTINUE

No .... 2 TERMINATE AND TALLY

3. Do you live in zipcodes, 60646, 60202, 60076 or 60077?

YES .... 1 CONTINUE

NO .... 2 TERMINATE AND TALLY

4. Do you or any member of your family work for an advertising agency, a market research firm or a company which makes, sells or distributes appliances?

YES .... 1 TERMINATE AND TALLY

NO .... 2 CONTINUE

5. When, if ever, did you last attend a market research group?

____

(IF LESS THAN 6 MONTHS AGO—TERMINATE AND TALLY)

Source: Chicagoland Field Service, 6200 N. Hiawatha, Chicago, Ill. 60646. Used by permission.

Having a relatively homogeneous group can help to a certain extent in maintaining group cohesiveness and getting everybody to participate. Nevertheless, a group in which the personalities of all participants are mutually compatible is more of an exception than the rule. Focus groups usually have one or two people who strive to dominate the discussion and one or two shy individuals who are reluctant to participate. Therefore a moderator must have good *observation, interpersonal,* and *communication skills* to recognize and overcome threats to a healthy group discussion. Exhibit 7.2 describes several skills a good moderator should have. As the descriptions of the skill dimensions indicate, a moderator should maintain a delicate balance on each; swinging too far toward an extreme on any one of them is likely to be detrimental.

In addition to the skills just described, moderators should have *interpretative skills.* Good moderators must be able to go beyond what the participants are saying by making note of such things as tone of voice, facial expressions, and other nonverbal behaviors of the group members. Such additional information will be valuable in interpreting the content of a focus group interview.[21] The importance of interpretative skills is evidenced in the following claim made by a research firm concerning its discussion leaders: "Leaders who have the knack of stimulating active participation in the discussions. Who benefit clients with insightful interpretations of what is said. *What is not said. What is meant*"[22] (emphasis added). Indeed, some have even suggested that a focus group is as much an observation technique as it is an interviewing technique.[23]

There are apparently no clear-cut guidelines regarding the physical or demographic characteristics of the moderator vis-à-vis those of the focus group participants. For instance, there is no agreement among researchers about whether the moderator should be of the same sex as the participants.[24] The ability to quickly establish rapport with the participants and make them feel at ease within the group seems to be more critical than the moderator's physical characteristics.

To summarize, not every individual, even one who may be good at one-on-one interviewing, may be an effective focus group moderator. Being able to conduct a focus group interview productively requires certain inherent personal skills as well as extensive training. Several commercial marketing research firms have ongoing, real-life training programs for their moderators. These programs involve conducting real-life focus groups—not for any particular client but purely for the sake of training moderators.[25]

---

[21] See, for example, "Focus Group Moderators Should Be Well Versed in Interpretative Skills," *Marketing News,* February 18, 1981, p. 23.

[22] "Marketing Research," a brochure describing the services offered by Elrick and Lavidge, Inc., 10 South Riverside Plaza, Chicago, Ill. 60606.

[23] Robert W. Joselyn, *Designing the Marketing Research Project* (New York: Petrocelli/Charter, 1977), p. 90.

[24] Bellenger, Bernhardt, and Goldstucker, *Qualitative Research in Marketing,* p. 16.

Conducting Focus Groups

Focus group sessions typically last one and a half to two hours. To use this relatively short period of time productively, the moderator must make the participants feel relaxed and comfortable with one another as quickly as possible. A skilled moderator can aid in creating an atmosphere conducive to a congenial and effective group discussion. Other means of attempting to generate a relaxed atmosphere include making the focus group room and its furnishings informal, and serving some light refreshments before the session starts, which are available throughout the session.

Focus group interviews are invariably recorded at least on audiocassettes for subsequent replay, transcription, and analysis. They can also be recorded on videotapes, especially if conducted in a well-equipped focus group facility rather than in an improvised setting such as a hotel room. Videotaping a focus group interview in the latter case can sharply increase costs.

Recording focus group interviews, in addition to increasing costs, has another potential drawback: The recording equipment, especially if it is perceived by participants as intrusive and threatening, can make creating a relaxed atmosphere difficult. A similar limitation is also associated with the presence of one-way mirrors, which most focus group rooms have. These mirrors enable client personnel to observe the group session without themselves being observed.

One way to minimize the detrimental effects of electronic devices or other means of observing focus groups is to make the equipment as inconspicuous as possible. Indeed, one research firm boasts: "Our mirrors are decorative walls . . . not tell-tale windows."[26] However, such measures still may not guarantee there will be no ill effects. This point is emphatically made by two practitioners who have conducted numerous focus groups.

> Researchers who conduct [focus] groups today often bring respondents into unfamiliar surroundings. And then they hope the "guinea pigs" won't be aware of, or inhibited by, the "laboratory" aspects of the facility. These

---

[25] For one such example, see Peggy Lang, "Women Executives Discuss Dressing the Part in Focus Group Sessions," *Marketing Today* (published by Elrick and Lavidge, Inc.), 19 (1) (1981), p. 1.

[26] Answers to Your Questions About Winona, Inc.," a brochure prepared by Winona, Inc., 8200 Humboldt Avenue So., Minneapolis, Minn. 55431, p. 12.

Exhibit 7.2 *Desirable Focus Group Moderator Skills*

- *Kind but firm.* In order to elicit necessary interaction, the moderator must combine a disciplined detachment with understanding empathy. To achieve this, the moderator must display a kindly, permissive attitude toward the participants, encouraging them to feel at ease in the group interview environment, while simultaneously insisting that the discussion remain germane to the problem at hand. Only with experience can the moderator achieve an appropriate blending of these two apparently antithetical roles.

  It is also the moderator's responsibility to encourage the emergence of leadership from within the group, while at the same time avoiding tendencies of domination of the group by a single member. The kindly but firm moderator must be sensitive to bids for attention and must maintain leadership without threatening or destroying the interactional process.

- *Permissiveness.* While an atmosphere of permissiveness is desirable, the moderator must be at all times alert to indications that the group atmosphere of cordiality is disintegrating. Before permissiveness leads to chaos, the moderator must reestablish the group purpose and maintain its orientation to the subject.

  The moderator must be ready and willing to pursue clues to information that may at first appear tangential to the subject; it may open new areas of exploration. The moderator must also be prepared to cope with expressions of unusual opinions and eruptions of personality clashes within the group. The manner in which these are handled may well be the difference between a productive and an unproductive group session.

- *Involvement.* Since a principal reason for the group interview is to expose feelings and to obtain reactions indicative of deeper feelings, the moderator must encourage and stimulate intensive personal involvement. If the moderator is unable to immerse himself or herself completely in the topic being discussed, the group will sense the detachment, and the depth contribution of the interview will be lost.

- *Incomplete understanding.* A most useful skill of the group moderator is the ability to convey lack of complete understanding of the information being presented. Although the moderator may understand what the participant is trying to express, by carefully inserting noncommittal remarks, phrased in questioning tones, the respondent is encouraged to delve more deeply into the sources of his or her opinion. The respondent is, by this process, able to reveal and elaborate on the kinds of information for which the group interview is designed. The goal is to encourage respondents to be more specific about generalized comments made by group members.

  The usefulness of this technique can be endangered if its application is inappropriate. If the "incomplete understanding" is a superficially assumed role, the group will soon detect this artificiality, and will feel that the moderator is playing some sort of cryptic game with the group. The group interview will then deteriorate into a sterile collection of mutual suspicions. Incomplete understanding on the part of the moderator must be a genuine curiosity about the deeper sources of the participant's understanding.

- *Encouragement.* Although the dynamics of the group situation facilitate the participation of all members in the interaction, there may be individuals who resist contributing. The skillful moderator should be aware of unresponsive members and try to break down their reserve and encourage their involvement.

  The unresponsive member offers a real challenge to the group moderator. There are numerous ways in which a resistant or bashful member can be encouraged to participate, such as by assigning the member a task to perform, or by providing an opening for his or her remarks. If this is inappropriately attempted, it may only reinforce a reluctance to participate in a verbal fashion. The ability to interpret nonverbal clues may provide a means of discovering a tactic to broaden the scope of the group's active participation.

- *Flexibility.* The moderator should be equipped prior to the session with a topic outline of the subject matter to be covered. By committing the topics to memory before the interview, the moderator may use the outline only as a reminder of content areas omitted or covered incompletely.

  If a topic outline is followed minutely, the progress of the interview will be uneven and artificial, jumping from topic to topic without careful transitions. This procedure communicates a lack of concern to the participants, for its mechanical nature makes the moderator appear to lack genuine interest in the responses.

  At the same time, the interview cannot be allowed to wander aimlessly. Under such conditions, control of the situation soon passes from the moderator to a self-appointed group leader.

  The group interview should be conducted the way one walks across a rope bridge. The handrails are gripped firmly and the objective is kept in mind constantly. If the bottom foot rope should break, the walk is continued hand over hand until the destination is reached. This requires an ability to improvise and alter predetermined plans amid the distractions of the group process.

- *Sensitivity.* The moderator must be able to identify, as the group interview progresses, the informational level on which it is being conducted, and determine if it is appropriate for the subject under discussion. Sensitive areas will frequently produce superficial rather than depth responses. Depth is achieved when there is a substantial amount of emotional responses, as opposed to intellectual information. Indications of depth are provided when participants begin to indicate how they feel about the subject, rather than what they think about it.

Source: Adapted from Danny N. Bellenger, Kenneth L. Bernhardt, and Jac L. Goldstucker, *Qualitative Research in Marketing* (Chicago: American Marketing Association, 1976), pp. 12–16; as adapted by the authors from Donald A. Chase, "The Intensive Group Interview in Marketing," *MRA Viewpoints* (1973). Used by permission.

> aspects include corporate marketers lurking behind a giant one-way mirror; microphones carefully displayed to avoid distraction; and wires and cables running behind dropped ceilings to avoid detection. . . . Our tests and observations have shown that most group respondents are *not* fooled in the slightest by the mirror. In fact, many instantly were inhibited, apprehensive, or suspicious.[27]

We do not mean to say that the use of recording devices and observation of focus groups by anyone other than the moderator must be avoided. In fact, a primary strength of a focus group interview is that it offers decision makers a direct report of what consumers say and how they feel about products, services, and the like. Furthermore, with current technology client personnel who are geographically dispersed may observe a group session in progress at a central location[28]—a valuable benefit that would be lost by a self-imposed ban on the use of equipment by which outsiders can watch a focus group session.

How, then, can one reconcile the need for a natural, relaxed atmosphere on the one hand and the decision makers' need to get the most out of a focus group on the other? A good compromise is to use whatever devices are essential to obtain maximum benefits from a group session but be open with participants about them. One should inform the participants about any recordings or observations that will take place during the session and explain the reasons for them. From the author's own experience with several focus groups, such candor, coupled with assurances from a well-trained moderator that no ulterior motives are involved, can put respondents at ease rapidly. In some instances it may be feasible and desirable to introduce client observers to the participants and even seat them in the focus group room.[29]

Advantages of Focus Groups

A number of articles have discussed many specific advantages of focus groups.[30] Rather than repeat any lengthy list here, we provide a synthesis in the form of *general advantages,* each of which encompasses several specific advantages cited in the literature.

---

[27] John O. Davies III and Jay F. Wilson, "Focus Groups Need New Focus, Changes in Their Facilities, Moderator Strategy, Marketing," *Marketing News,* May 13, 1983, Section 2, p. 19.

[28] For an illustration of this capability, see F. George du Pont, "Use Telephone to 'Attend' Distant Qualitative Research Sessions," *Marketing News,* May 13, 1983, Section 2, p. 19.

[29] The use of such an approach is reported by Davies and Wilson, "Focus Groups Need New Focus."

[30] See, for example, Bellenger, Bernhardt, and Goldstucker, *Qualitative Research in Marketing,* pp. 17–28; John M. Hess, "Group Interviewing," in *New Science of Planning,* ed R. L. King (Chicago: American Marketing Association, 1968), pp. 193–196; Alfred E. Goldman, "The Group Depth Interview," *Journal of Marketing,* 26 (July 1962), pp. 61–68.

Richness of Data. A well-conducted focus group interview can provide numerous important insights that a series of one-on-one depth interviews may be unable to generate.[31] Respondents are generally less likely to feel inhibited in the company of others similar to them than they are when alone with an interviewer. They are also likely to be more creative in terms of offering insights about the topic being discussed. A comment made by one participant may trigger a stream of new ideas from the others; this reaction obviously cannot occur in a series of individual interviews.

Versatility. Focus groups can be used to gain insight into a great variety of problems (examples of common applications of focus groups will be described in a subsequent section). Almost any product, service, concept, issue, or the like, can be productively discussed in a focus group setting. Focus groups are also versatile in another respect: A number of different techniques (e.g., a projective question or a TAT type of picture) and stimuli (e.g., a product or an ad) can be used in conjunction with or during a focus group in order to increase the productivity of the discussion.[32] As one research firm puts it: "Consumers [in our focus groups] have turned knobs, opened doors, used pots and pans, eaten foods, ripped and pulled at materials, laughed and frowned at greeting cards, and in hundreds of other ways have acted naturally, as consumers normally do when confronted with a new idea or a new product designed for their consumption."[33]

Ability to Study Special Respondents. Focus groups may be the only feasible method of gathering data in certain situations where conducting one-on-one interviews is unproductive or impossible. Such situations involve special types of respondents such as children, especially very young children. Interviewing children on a one-on-one basis is often difficult and frustrating. A researcher using focus group interviews for research with children is likely to be more successful in generating useful information. According to Child Research Services (CRS), a division of McCollum/Spielman specializing in studying children: "While there may be major impediments to talking candidly with children, this does not mean that they are uncommunicative. In fact, children communicate very well through play, which has become the linchpin of CRS's methodology for extracting information from small children in *focus groups*"[34] (emphasis added).

---

[31] Although this view is held by a majority of researchers, it is not unanimously believed; for some differing viewpoints, see, for example, Fern, "The Use of Focus Groups for Idea Generation"; Melanie S. Payne, "Individual In-Depth Interviews Can Provide More Details Than Groups," *Marketing Today,* 20 (1) (1982), pp. 1–3.

[32] Goldman, in "The Group Depth Interview," discusses the use of several techniques and stimuli during focus group sessions.

[33] "Group Interviewing," Burke Marketing Research, Inc., p. 1.

Other groups of hard-to-interview respondents include professionals such as doctors and lawyers who frequently refuse to be interviewed individually. However, these professionals may be receptive to a group interview, which gives them an opportunity to be with their colleagues and compare notes with them on topics of mutual professional interest.[35]

Impact on Managers. As has been pointed out in previous chapters, one of the pet peeves of managers (i.e., users of marketing research) is that researchers overemphasize sophisticated techniques and statistical results. Consequently, quantitative-research reports often do no more than collect dust on managers' shelves. However, qualitative-research studies in general, and focus group studies in particular, stand a much better chance of making an impression on managers (i.e., of being acted upon). Why? Because managers have the opportunity to get actively involved in and observe a focus group study instead of merely reading a research report. Moreover, key ideas emerging from a focus group, and a few crucial statements made by the participants, can sometimes have a greater impact on decision makers than a report full of impressive tables and statistics. Relevant quotes from a focus group interview may be worth a thousand numbers, so to speak. The following example is a case in point from the author's personal experience.

> **Example.** A securities brokerage firm commissioned a series of focus groups to understand customers' perceptions of and satisfaction with brokers who handled their transactions. One respondent in a male focus group made the following statement, indirectly describing his own frustration with brokers: "Sometimes, consumers tend to go along with what the broker says because they don't want to sound like 'morons'!" All the other participants nodded in agreement. Just as if designed to reinforce this feeling, another male respondent in a *different* focus group held in a *different* city described an ideal broker this way: "I like to be able to ask dumb questions. How he [the broker] reacts to the dumb question is very important to me." The two statements, emerging from two different focus groups, conveyed the same powerful message: Brokers tend to talk over customers' heads and make them feel stupid and uncomfortable. Needless to say, these customer remarks and others like them made a deep impression on the firm's executives.

---

[34] "The Child Research Services," a brochure published by McCollum/Spielman Associates, Inc., 18 East 48th Street, New York, N.Y. 10017.

[35] Lee Adler, "To Learn What's on the Consumer's Mind, Try Some Focused Group Interviews," *Sales & Marketing Management,* 122 (April 9, 1979), pp. 76–77.

The point here is that "colorful" and natural customer statements, capable of having a significant impact on decision makers, can only emerge from a focus group setting.

Disadvantages of Focus Groups

As in the case of advantages, a concise set of general or global disadvantages—synthesizing numerous specific drawbacks cited in the literature—is presented here.

Lack of Generalizability. To emphasize what we have already seen, *focus group results cannot be viewed as conclusive.* The small sample sizes and homogeneous group compositions, while crucial to conducting focus groups effectively, make any insights gained from them highly tentative. Moreover, a person who agrees to participate in a focus group interview is typically more outgoing than the average consumer. This bias, coupled with the possible presence of professional respondents who routinely participate in numerous group interviews, can further erode the representativeness of a focus group.

Opportunity for Misuse. Some of the very strengths of focus groups—richness of data and impact on managers, for example—can turn into serious drawbacks when the technique is employed by careless researchers and managers. One potential misuse of focus groups occurs when managers yield to a temptation to generalize a few key remarks made by participants, especially when similar remarks are made in several focus groups. Generalization is a serious mistake, since the primary purpose of focus groups is simply to generate insights or hypotheses, which must be verified through more formal research using representative samples.

Another potential misuse stems from the opportunity for moderators and managers to interpret focus group data any way they want to. Although this problem can occur with the use of any qualitative-research technique, it is particularly serious in focus group interviews. Since focus groups are informal, and since the data they generate are typically varied and rich, moderators and managers with preconceived notions about an issue may be able to find support for their positions by interpreting the data selectively—that is, by hearing and emphasizing only those remarks that do not dispute their opinions. If such selective perceptions occur, the focus group results will be questionable. Objective moderators and open-minded managers are essential to ensure proper use of focus groups. The former is especially critical: A moderator who is not objective can misinterpret the focus group data and also lead the discussion in a direction consistent with his or her personal viewpoints.

Cost. A focus group interview that is professionally conducted in a research firm's facility can easily run up a total bill of around $2500. If this figure sounds high, consider the following variety of expensive items essential for a well-conducted focus group: Numerous telephone calls are usually needed to recruit even a handful of qualified participants—several research firms charge around $20 to $25 just to recruit *each* respondent (this cost will be even higher if rare or hard-to-reach respondents are to be recruited). And, normally, the firm must recruit a few extra respondents because of the possibility of no-shows at the time of the interview. In a typical consumer focus group an incentive of around $35 is usually paid to each recruit who shows up (even to those who have to be turned away if everybody recruited shows up). In special focus groups involving hard-to-get respondents (e.g., senior business executives, doctors) the incentive per respondent is usually much higher—in the $75-to-$100 range. A qualified moderator's fee for conducting a focus group may run around $350. Adding to the above items such things as renting the focus group facility, transcribing and interpreting the interview, and videotaping (each of these usually cost several hundred dollars apiece) shows why a single focus group can cost $2500.

Some managers, though, may perceive the total cost of even a series of focus groups as low because a typical quantitative project conducted by a research firm can carry a price tag of $15,000 to $20,000. However, this perception is fallacious since focus groups are a prelude to, and *not* a replacement for, quantitative research. Furthermore, on a *cost-per-respondent* basis, focus groups are extremely expensive.[36] Therefore focus groups that are conducted without the managers' establishing a need for them, and without thinking through what they can and cannot deliver, will be an unnecessary drain on a firm's resources. Managers who rely on focus group research, believing that the technique is a quick substitute for formal research, run the risk of making bad decisions in addition to wasting valuable resources.

Focus Group Applications

Focus groups can be used in almost any situation requiring some preliminary insights (i.e., exploratory research). However, specific applications of focus groups are too numerous to discuss here. Therefore this section simply presents three broad application categories that include the most frequent applications of the technique: understanding consumers, product planning, and advertising.

Understanding Consumers. Focus groups can be helpful in attempting to get a feeling for consumers' perceptions, opinions, and behavior concerning products or services. For example, a manufacturer of cake mixes may be interested in questions such as these: What do consumers like about baking? What do they dislike? Why do they bake? How do they bake? What words/terms do they use in describing baking products and their use? Insights about these issues can be useful in a variety of situations, such as when the firm wants to identify marketing problems or opportunities worthy of further investigation, or when it is planning a formal consumer survey and wants to know what questions are worth including in it and how they should be worded.

---

[36] These points are emphasized by Adler, "To Learn What's on the Consumer's Mind."

Product Planning. Focus groups are useful in generating ideas for new products. For instance, consider the case of Glass*Plus, an all-purpose, light-duty spray cleaner marketed by Texize Chemicals. Before conception of the idea for Glass*Plus, Texize had conducted a survey of consumers concerning their use of its products, one of which was Fantastik, a heavy-duty cleaner. This survey showed that many consumers were using Fantastik to clean glass despite the product's not being recommended for cleaning glass. Texize therefore conducted a series of focus group interviews to understand why consumers were using Fantastik to clean glass when products like Windex were specifically available for that purpose. Insights from these focus groups pointed out the need for an intermediate-strength product that could be used for cleaning glass as well as for other household-cleaning chores. The idea for Glass*Plus was thus born.[37] Focus group interviews can also be used to suggest ideas for modifying existing products to improve their market performance.

Advertising. Perhaps the most frequent application of focus groups is in developing creative concepts and copy material for advertisements. Advertising agencies almost constantly face the challenge of changing advertisements to give them a new look and to increase their impact on consumers. They are heavy users of focus groups since a major strength of the technique is its ability to generate new ideas. Focus groups can suggest appropriate words, slogans, copy themes, pictures, and product benefits for use in advertisements. They can even offer hints on acting for television commercial models, as the following example demonstrates.

> **Example.** A marketer of a painkiller specifically intended for people with a certain type of disease wanted to develop an effective television commercial for it. A group of people afflicted by the disease was assembled, and each person was asked to describe and discuss the pain resulting from the disease. The group discussion was videotaped. "By closely examining the videotapes, [it was] found that there were a series of some 45 nonverbal facial and body expressions people used

---

[37] Bellenger, Bernhardt, and Goldstucker, *Qualitative Research in Marketing*, p. 22, discuss this and other examples in greater detail.

Research in Use 7.1 *The Johnson Company*

Focus Group Application

The Johnson company developed a new filter to be used in car air conditioning systems. Management wanted to find out the feasibility of the new product and develop a workable marketing plan. A two-stage research process was followed. Focus group interviews were used in stage 1 to help develop hypotheses to identify potential markets, to determine advantages and disadvantages of the product from the consumer viewpoint, and to identify specific points for questionnaire design.

The focus group interviews indicated that families in which one or more members had allergy or respiratory problems might be the best prospects for the new product. Persons seriously concerned about air pollution were also identified as good potential buyers. The major disadvantages of the product were the performance capability of the filter and the cost of replacement cartridges. Some individuals feared that the filter would cause their car's air conditioning system to malfunction. Nonallergic consumers expressed doubt about their need for the product. An unexpected resistance occurred when consumers were informed that the filter would need to be changed periodically.

After hearing the results of the focus group, the client wanted to proceed immediately with market introduction as a result of the findings that seemed favorable. On the advice of the research firm, the quantitative study in stage 2 of the research process was conducted. An analysis of 1500 respondents in five cities showed that the original marketing strategy for introducing the new product was not economically feasible. This led to the development of an alternative marketing plan.

Source: Reprinted from Keith K. Cox, James B. Higginbotham, and John Burton, "Applications of Focus Group Interviews in Marketing," *Journal of Marketing*, 40 (January 1976), pp. 78–79. Used by permission of the publisher, the American Marketing Association, Chicago, Ill.

when they spoke about pain and its relief—gestures which were incorporated into the commercials used to advertise the product."[38]

To further demonstrate the application of focus groups and to emphasize their proper use, we provide a real-life illustration in Research in Use 7.1. In this illustration, note especially that the focus groups were helpful in gaining an initial understanding of the target market and in developing an appropriate questionnaire for the quantitative study; that the company executives were eager to take final action simply on the basis of *some* of the focus group results (this temptation is not uncommon); and that had such final action been taken, the consequences could have been very unfavorable.

---

[38] Lisa Gubernick, "Are Market Research Groups Out of Focus?" *Adweek*, September 5, 1983, p. 22; see "Behind the Scenes at an American Express Commercial," *Business Week*, May 20, 1985, pp. 84–88, for other applications of focus groups in the area of advertising.

OTHER QUALITATIVE-RESEARCH TECHNIQUES

The traditional focus group interview as described in the preceding sections is the most frequently used, but not the *only*, form of qualitative research. Several techniques presented in the previous chapter as illustrations of nonstructured data collection (e.g., in-depth interviews and projective techniques) are also types of qualitative research, but they will not be discussed again here. Rather, techniques we have not seen thus far, several of which are variations of the traditional focus group interview, will be presented to illustrate the wide variety of choices available in qualitative research.

Dual-Moderator Group

Dual-Moderator Group
A type of focus group in which there are two moderators, one responsible for ensuring the smooth flow of the session and the other responsible for ensuring that key issues get discussed.

The ***dual-moderator group,*** as its name implies, is a focus group interview conducted by two moderators. Dual-moderating can be particularly useful in focus groups involving discussion of a highly technical topic.[39] An example is a group of electronic engineers discussing new concepts for computer memory chips. If dual-moderating is used in such a group, the two moderators will have distinct roles: One will be responsible for the smooth flow of the session, and the other will be responsible for ensuring discussion of specific technical issues that might otherwise be overlooked. The former is a sort of social moderator, while the latter is a technical moderator. To be truly beneficial, dual-moderating requires effective coordination between the two moderators as well as good rapport between each moderator and the respondents. This coordination may not always be easy to accomplish.

Dueling-Moderator Group
A variation of the dual-moderator focus group. The two moderators in a dueling-moderator group deliberately take opposite positions on the discussion topic so as to stimulate a lively and potentially revealing exchange of views among the participants.

An interesting variation of the dual-moderator group is the ***dueling-moderator group*** in which the two moderators deliberately take opposite positions on the issue to be discussed. The moderators get together before the focus group session starts and agree to a plan outlining what positions each will support, as well as how often and when they will do so during the session. Dueling-moderator groups are appropriate for examining and understanding people's perceptions about controversial social or political issues.[40] Properly conducted dueling-moderator focus groups can offer valuable suggestions for correcting misperceptions and changing opinions. As you might expect, however, such focus groups conducted by ill-qualified and inexperienced moderators run the risk of resulting in pandemonium rather than providing useful insights.

---

[39] J. Pope, "Six Ways to Improve Today's Focus Group Research," *Advertising Age,* 48 (July 11, 1978), p. 150.

[40] Buncher, "Focus Groups Seem Easy," elaborates on these points.

Other extensions of the dual-moderator group include the following:

Respondent-Moderator Group *A focus group interview in which the moderator may ask selected participants to play the role of moderator to improve group dynamics and productivity.*

Client-Participant Group *A focus group interview in which client personnel are introduced to respondents and made part of the group.*

- ***Respondent-moderator groups*** in which the real moderator may ask selected participants to temporarily play the role of moderator to improve group dynamics and productivity.
- ***Client-participant groups*** in which client personnel are introduced to respondents and made part of the group. The client-participant's primary role (in addition to gaining firsthand knowledge about the session) is to offer clarifications and other relevant inputs that can improve the efficiency and effectiveness of the discussion.[41]

Groups Receiving Instant Feedback About Participants' Collective Opinions

Consensor *An electronic device used in certain focus groups. It consists of a set of hand-held terminals through which participants can express the nature and strength of their feelings about any issue raised by the moderator. The feelings of the group are instantly tabulated and displayed as a histogram on a video screen.*

Some research firms are now using electronic technology to cut down on unproductive discussion time and hence increase the value of focus groups. The electronic gadgetry used for this purpose is called the ***consensor,*** and it provides each participant with a hand-held terminal. All terminals are connected to a common video-display screen visible to the entire group. Each hand-held terminal has two dials: an opinion dial, which is used to anonymously express an opinion about an issue on a scale of 0 (no) to 10 (yes), and a weighting dial, which is used to express strength of an opinion on a 0%–100% scale.

The consensor is a valuable timesaver when the moderator wants to poll the participants on some issue germane to the group discussion. Participants simply express their feelings using their hand-held terminals, and the results are instantly tabulated and displayed in the form of a histogram on the video screen.

According to one practitioner who has used this technology: "The Consensor allows a group leader to focus more on the issue of *why* people feel the way they do, and not *how* they feel."[42] The consensor provides a tangible stimulus (i.e., the video display) to guide the discussion, and, apparently, it also stimulates greater participant interest and involvement than a traditional focus group does. Using the consensor does involve a potentially serious risk: The histograms and numbers may tempt clients to believe that the results are *generalizable.* Such a belief constitutes a misuse of focus groups.

---

[41] Further discussion of these two and other focus group variations are found in Buncher, "Focus Groups Seem Easy,"; Pope, "Six Ways to Improve."

[42] Lynn M. Moss, "Consumer's 'Visual Feedback' of Opinions Could Revolutionize Focus Group Research," *Marketing News,* November 26, 1982, Section 2, p. 6.

Crowded One-on-One Interview

Crowded One-on-One Interview *A variation of the depth interview, conducted by a professional interviewer in the presence of up to three client personnel, who themselves can probe the respondent toward the end of the interview.*

A ***crowded one-on-one interview*** is actually a variation of the in-depth interview. In a crowded one-on-one interview up to three client personnel are present in the room and observe a depth interview, as it is conducted by a professional interviewer in a conventional fashion. After the interview is over, the client personnel can ask additional questions or seek clarifications of responses already given. According to one marketing researcher: "Crowded one-on-ones are particularly effective when new concepts, designs, advertising, promotional messages, etc. are in the formulation stage."[43]

Shopper Study

Shopper Study *A study in which a professional interviewer plays the role of a customer and gathers relevant data (through questioning as well as observation) on issues of interest.*

In a ***shopper study*** a professional interviewer plays the role of a customer and gathers relevant data (through questioning as well as observation) on issues of interest. Shopper studies that generate data in an unobtrusive or disguised fashion (to ensure data validity) can provide valuable insights. For instance, a retail establishment whose sales depend heavily on its floor salespeople can hire interviewer-shoppers to gather data about the selling techniques used by the salespeople. Interviewer-shoppers can also be hired to generate information about such things as competitors' prices, product assortments, and product display effectiveness. Shopper studies have very many such applications. Research in Use 7.2 describes Hospitality Standards, a firm that conducts shopper studies for the hotel industry.

The preceding list of qualitative-research techniques is by no means exhaustive. A researcher's imagination is the limit on the specific form qualitative research takes in a given situation, as demonstrated by the variations of traditional techniques illustrated in this section. Such flexibility and opportunity to improvise is a major strength of qualitative research. Unfortunately, this very strength is also an invitation for ill-qualified and/or unscrupulous researchers to misuse and abuse qualitative research.

A high degree of subjectivity is involved in every stage of the qualitative-research process, including the critical stages of sample selection, data collection, analysis, and interpretation, which has a major bearing on the validity of the results. Therefore using qualitative research carelessly and irresponsibly can give a very misleading picture of reality, on top of wasting a firm's resources. The three limitations discussed in connection with focus groups—unprojectable results, opportunity for mis-

---

[43] Richard E. Matheson, "Qualitative Research Methods (Other Than Focus Groups) Can Provide Valuable Information," *Marketing News*, May 13, 1983, Section 1, p. 14. This reference discusses crowded one-on-ones as well as other variations of in-depth interviews; it also mentions several potential applications of the shopper studies described next in this text.

Research in Use 7.2 *Hospitality Standards*

Application of Shopper Studies

With competition hot among luxury hotels, business is bubbling for lodging consultants, those inspectors who check out a place by checking in. Usually registering under assumed names, these company-paid snoops behave like eccentric guests, evaluating cleanliness, service, and staff competence.

One such consultant is June Michaelson, 57, whose firm, Hospitality Standards, will inspect 120 hotels this year, including those in the Marriott, Ritz-Carlton, and Westin chains. For $4,000 plus expenses, a two-person team spends three days examining every aspect of a hotel. Michaelson, who may change clothes as many as seven times a day to confuse the staff, typically clocks the registration clerk as he logs in guests, and calls room service for a fried egg sandwich at 1 a.m.

No area goes unexamined. "I pop into the men's room when no one's looking," she says. "Funny thing about hotel men's rooms: They're cleaner than the women's. Most general managers are men, so they pay a visit and know if it is in good shape or not."

In her business Michaelson has a special edge. From 1979 until 1982, she was editor-in-chief of the *Mobil Travel Guide.* Since she updated the criteria for its rating system—one star, "above average," to five stars, "one of the best in the country"—she is in a unique position to suggest how a hotel can boost its reputation. Says Michaelson: "Most of my clients want to better their service and some want that elusive five-star rank."

Source: David J. Morrow, "Hotel Harasser for Hire," *Fortune,* September 26, 1988, p. 16. Reprinted by permission.

use, and high cost per respondent—apply to other qualitative techniques as well and must be kept in mind by all qualitative researchers and potential users of the research results.

SUMMARY

Qualitative research involves generating information that cannot be summarized in the form of numbers. It is characterized by nonstructured, flexible data collection from relatively small respondent samples. The primary purpose of qualitative research is to generate insights and ideas. Therefore the proper place for qualitative techniques is in exploratory-research projects. In contrast, quantitative tehchniques are more appropriate for conclusive-research projects.

Qualitative research is no less important or valuable than quantitative research. The two types of research should be viewed as complementing each other—qualitative research can be a valuable, and many times necessary, prelude to quantitative research.

The most widely used qualitative-research technique is the focus group interview. Three major determinants of focus group effectiveness are (1) group composition, (2) moderator characteristics, and (3) the manner in which the focus group is conducted (i.e., group atmosphere). Although there are no absolute guidelines for these factors, several rules of thumb have emerged through experience.

Focus groups typically have 8 to 12 participants. Homogeneity of group members, especially on characteristics such as sex and age, is crucial. Focus group participants must also have had some experience with or exposure to the discussion topic, and they should not be professional respondents. A generally accepted practice is to conduct at least two focus groups on a given topic.

To be effective, a focus group moderator should have good observation, interpersonal, communication, and interpretative skills. The physical or demographic characteristics of the moderator are apparently not as critical as his or her ability to establish rapport with respondents and guide the discussion in a natural, yet focused, fashion.

To be of maximum benefit, focus groups should be conducted in a relaxed, comfortable atmosphere. Having the appropriate physical setting and amenities can be helpful. Focus groups are invariably recorded by electronic means and are often observed unobtrusively by client personnel. An important factor here is to ensure that the means used for recording and observing do not adversely affect group atmosphere and hence the quality of the discussion.

The major advantages of focus groups include the richness of the data they can generate, their versatility—in terms of their wide application potential as well as their flexibility to be used in conjunction with other techniques—their usefulness in gathering data from otherwise hard-to-interview respondents, and their ability to make a lasting impression on managers. A limitation of focus groups is that their results are not generalizable. Moreover, the subjectivity involved in conducting and interpreting focus groups is an invitation for irresponsible users to abuse the technique. Focus groups are also expensive, especially on a cost-per-respondent basis.

A general application of focus groups is to develop an understanding of consumers and their perceptions. Focus groups can reveal signs of potential problems, suggest reasons for existing problems, and point out possible marketing opportunities. They can also be helpful in designing formal-research studies of consumers. More specific applications of focus groups occur in the area of product planning and advertising.

Qualitative-research techniques other than the traditional focus group interview include some techniques that were discussed in previous chapters (e.g., case studies, in-depth interviews, and projective techniques); variations of focus group interviews (e.g., dual-moderator and dueling-moderator groups); and variations of in-depth interviews such as crowded one-on-ones and shopper studies. While the techniques can provide valuable insights, they also suffer from the potential limitations of focus groups.

KEY TERMS

Qualitative research
Quantitative research
Focus group interview
Professional respondent
Dual-moderator group
Dueling-moderator group
Respondent-moderator group
Client-participant group
Consensor
Crowded one-on-one interview
Shopper study

QUESTIONS

1. Summarize the key differences between qualitative and quantitative research.
2. "Qualitative research must not be conducted in situations where obtaining quantitative data is the ultimate objective." Discuss this statement.
3. Discuss the reasons why a focus group should consist of similar participants.
4. How many focus groups should be conducted on any given topic? Explain your answer.
5. Who are professional respondents? In what ways can they affect a focus group?
6. Briefly describe the skills that an effective focus group moderator should have.
7. "It is not advisable to record a focus group interview through electronic means." Discuss the pros and cons of this statement, indicate what *your* position is, and defend your position.
8. State the general advantages and disadvantages of focus groups, and describe each in a sentence or two.
9. Why are focus groups widely used in advertising?
10. What are the differences between dual-moderator and dueling-moderator focus groups?
11. State and briefly describe two variations of the traditional in-depth interview.
12. For each of the following scenarios, indicate whether qualitative or quantitative research is more appropriate. Also recommend a specific technique for each, and justify your answer.

a. A manufacturer of food coloring wants to know how often and for what purposes or occasions consumers use food-coloring products.

b. A firm marketing chewing gum has two alternative wrapper designs for the product and is wondering which one is likely to result in higher sales.

c. A medical products company is in the final stages of perfecting a new method of contraception. It wants to guage the reactions of men and women to the new method and their potential acceptance of the method.

d. A national charitable organization wants to identify ways of sharply increasing the amount of contributions from the public.

CHAPTER 8

Introduction to Experimentation in Marketing

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between descriptive and experimental research.*
- *State the three conditions that should be satisfied to infer with confidence that some variable has a causal influence on another.*
- *Distinguish between, and state the relative strengths and weaknesses of, laboratory and field experiments.*
- *Define and illustrate internal validity and external validity of experimental results.*
- *State and discuss the factors to be considered in choosing between laboratory and field experiments.*
- *Define and illustrate the following threats to the internal validity of experiments: history, maturation, pretesting, instrument variation, selection, and mortality.*
- *Define and illustrate the following threats to the external validity of experiments: reactive bias, pretest-manipulation interaction bias, and nonrepresentative-sample bias.*

Questions like the following are frequently faced by decision makers in a variety of organizations:

- Question facing the advertising manager of a firm marketing canned soups: "Will replacing commercial A with commercial B for our canned soups lead to a marked increase in consumer preference for our brand?"
- Question facing the marketing vice president of a retail store chain selling a line of fashion clothing: "Can we improve the profitability of our fashion clothing line by increasing its price by 10%?"
- Question facing the sales manager of an industrial marketing firm selling components to electronic-goods manufacturers: "Will an increase in the average number of sales calls per customer from six to eight per year significantly improve sales?"
- Question facing the manager of a supermarket: "Will decreasing the shelf space allocated to brand X detergent by 25% significantly lower its sales?"
- Question facing the fund-raising chairman of a charitable organization: "Will it be worthwhile to mail to last year's donors an attractive (but expensive) brochure describing our organization's activities and soliciting higher contributions for this year?"

Decision makers typically answer questions like these on the basis of results from limited exploratory or descriptive research (see Chapter 4 for a discussion of these types of research) or, sometimes, solely on their intuitions. Yet each of these illustrative questions implies that an *accurate* answer can only be ascertained through some form of *controlled* research.

That is, each question involves investigating the nature of a *causal association* between two things: the impact of type of commercial on consumer preference in the first case, the impact of a price increase on profitability in the second, the impact of a greater number of sales calls per customer on sales in the third, the impact of reducing shelf space on sales in the fourth, and the impact of using a brochure on the volume of contributions in the last case. As we saw in Chapter 4, given data pertaining to two variables, the assurance with which one can estimate the impact of one variable on the other critically depends on the conditions under which the data are gathered. Of all the various types of research that may suggest causality between two variables, the greatest assurance that a causal inference is sound will stem from experimental research.

Definition. An ***experiment*** is a procedure in which one (or sometimes more than one) causal variable is systematically manipulated and data on the effect variable are gathered, while controlling for other variables that may influence the effect variable.

This chapter will provide an overview of experiments in the field of marketing research, including the major approaches to conducting marketing experiments and potential biases that can limit the degree of researcher control in such experiments. The next chapter will cover certain standard experimental designs useful for marketers and discuss the advantages and disadvantages of each. To place marketing experiments in their proper perspective, we will first review the distinction between experimental and descriptive research and the general conditions that must be met for making conclusive statements about the nature of causality between variables. The next two sections provide this review.

DESCRIPTIVE VERSUS EXPERIMENTAL RESEARCH

Keith Cox and Ben Enis succinctly describe the distinguishing features of experimental research as follows (in the quotation "independent variable" and "experimental variable" are what we referred to earlier as "causal variable," and "dependent variable" is what we referred to as "effect variable"):

> Experimentation differs from alternative methods of marketing research in that in experimentation the researcher manipulates the independent variable or variables before measuring the effect upon the dependent variable. For example, the effect of price changes on sales volume of a particular product can be examined by actually varying the price of the product. A nonexperimental approach would be to ask consumers whether they would buy more of the product if [its price were lowered]. The manipulation of independent variables, together with procedures of controlling extraneous variation . . . forms the basis of the power of experimental

research relative to other research techniques. The better the researcher's control over the experimental variables and extraneous variation, the more confident the researcher can be that he is in fact determining cause and effect relationships.[1]

The distinction between descriptive and experimental research is more a matter of degree than of kind. While descriptive-survey data will merely suggest causation, data generated through experimental research will increase the degree of confidence one can have in any suggested cause. Although, theoretically, a completely controlled experiment can indicate for sure whether something is caused by something else, in marketing practice complete control is rarely possible. Therefore seldom can causation be conclusively established in practical settings.[2] This point is especially noteworthy because the term *marketing experiment* does not necessarily imply that evidence of association between variables based on experimental data is *proof* that the association is a causal one. The discussion in the next section should further reinforce our contention that experimental research is not necessarily infallible.

CONDITIONS FOR INFERRING CAUSALITY

When can one confidently conclude that some variable (say X) has a causal influence on another variable (say Y)? The literature on experimental research suggests three rather intuitive conditions that must be satisfied for making a statement such as "If X, then Y":[3]

1. Temporal ordering of variables. The variable X (or a change in X) must occur before the variable Y (or a change in Y) is observed.
2. Evidence of association. The data on variables X and Y must suggest that the two are related in some fashion. In Chapter 4 we saw briefly how evidence of association (or lack thereof) between variables can be ascertained through cross tabulation. This technique and other analysis procedures for examining associations will be covered in detail in Chapters 16 and 18.
3. Control of other causal factors. Unless all potential causal factors other than X are satisfactorily controlled or accounted for, the statement "If X, then Y" may be false *even when* X precedes Y and evidence of association between the two exists.

---

[1] Keith K. Cox and Ben M. Enis, *Experimentation for Marketing Decisions* (Scranton, Pa.: International Textbook, 1969), p. 5.

[2] For a compelling argument, see Fred N. Kerlinger, *Foundations of Behavioral Research* (New York: Holt, Rinehart and Winston, 1973), p. 380.

[3] See, for example, Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook, *Research Methods in Social Relations* (New York: Holt, Rinehart and Winston, 1959), pp. 83–88.

All three conditions must be met before causality between X and Y can be conclusively established. Of the three, the third condition is most critical and perhaps also the most difficult to satisfy in practical research projects. Proper temporal ordering of variables and evidence of association can usually be inferred from observation and/or questionnaire data collected through descriptive research. But some form of experimental research is invariably required to account for other causal factors.[4]

Even in experimental settings, however, one cannot always be sure that *all* extraneous factors have been accounted for. Sometimes experimental researchers may not even be aware of some potential causal factors. As Charles Ramond correctly observes:

> Designing a conclusive marketing experiment requires a thorough knowledge of the specific situation and the ability to control, observe, or even foresee enough of the possible causes of the effect under study to permit observation of the undiluted influence of some of them. This ability can sometimes be learned from a thorough analysis of the sales and marketing history of the product involved. . . . In other cases there may be no way of knowing until some experiments have been done.[5]

Fortunately, most real-life situations call for *reasonable* (rather than *complete*) assurance of control, which sound experimental studies are capable of providing. Nevertheless, one must be cautious in making causal inferences simply because the first two conditions are satisfied.

To highlight the fallacy of such inferences, consider the well-known fact that roosters crow every morning before the sun rises. In this case roosters' crows certainly precede sunrises, and there is evidence of strong association between the two. Yet it is obviously ridiculous to infer that the sun rises *because* roosters crow. Although that one would make such an inference is unthinkable, the moral of this example is often subtly ignored by those who make causal inferences solely on the basis of proper temporal sequencing and evidence of association. The following example illustrates this point.

Example. In a *Time* magazine article examining the MTV cable channel devoted to music videos, an executive associated with MTV is quoted as saying: "I think Duran Duran (a British rock music band) owes its life to MTV." In apparent support of this executive's inference, the article states:

> Even the record industry . . . noticed that the Duran Duran album, *Rio*, was being sold out at half the record stores in Dallas and was gathering dust in the other half. A check of the local television listings showed that

---

[4] Several real-life studies leading to incorrect causal inferences owing to inadequate control of extraneous factors are described by Charles Ramond, *The Art of Using Science in Marketing* (New York: Harper & Row, 1974), pp. 180–184.

[5] Ramond, *Art of Using Science*, pp. 187–188.

> parts of the city that were wired for cable and carrying MTV were the very same parts where the album was flourishing.[6]

The implication here is that availability of the MTV channel *caused* a sharp increase in sales of the Duran Duran album. The former certainly preceded the latter, and there is no question that the two were strongly associated. Yet, borrowing some terminology from the criminal justice system, the evidence here is mainly "circumstantial" and does not show "beyond a reasonable doubt" that MTV was responsible for the sales success of Duran Duran's album in some parts of Dallas. Other uncontrolled factors might have been responsible for what was observed. For instance, availability of the MTV channel and high sales of the album may *both* have been caused independently by unobserved factors such as the socioeconomic status and interests of residents in various parts of Dallas. If so, the observed association between availability of the MTV channels and the record album's sales is spurious. This spurious association will vanish if the extraneous factors are explicitly incorporated into the analysis.

LABORATORY VERSUS FIELD EXPERIMENTS

Consider the following scenario, which we will use to illustrate our discussion in this and subsequent sections of the chapter: The advertising agency of Gourmet Food Products Company (GFPC) has developed two quite different television commercials (say A and B) for the company's line of frozen pizzas. The final decision about which commercial to use in an upcoming national campaign rests with GFPC's advertising manager, Mr. Thompson. Before choosing one of the two commercials, Mr. Thompson wants to get a feeling for the potential impact of each on consumer preference for GFPC's frozen pizzas. He therefore asks the ad agency to conduct an appropriate marketing experiment to assess the differential impact of one commercial over the other. As summarized next, the ad agency can use one of two general experimental approaches for generating the information requested by Mr. Thompson.

> **Approach 1.** Invite a group of, say, 100 consumers to the agency's marketing research facility. Divide the group randomly into two similar groups of 50 consumers and take each to a different television-viewing room. Show each group an hour-long episode of a current television program with a normal number and variety of commercials inserted into it at appropriate time slots. Manipulate the set of commercials to be seen by the two groups such that one group will see test commercial A twice during the program and the other group will see test commercial B. In other words, both groups will see the same

---

[6] "Sing a Song of Seeing," *Time*, December 26, 1983, p. 56.

program and set of commercials except for the differences in the test commercials. To prevent participant knowledge of the experimental manipulation, and to evoke normal reactions to the test commercials, disguise the purpose of the study by telling consumers that the study's purpose is to get their reactions to the television program. After the consumers have seen the program, administer a questionnaire to obtain their reactions to the program and to the various products advertised during the program, including GFPC's frozen pizzas. Compare the mean preferences for GFPC's frozen pizzas for the two groups to assess the impacts of commercial A and commercial B.

Approach 2. Select two test cities that are as similar as possible in terms of consumer demographics and food purchase characteristics. Select an appropriate television program that will be shown simultaneously in both cities but one that allows for airing different local television spots in the two cities during the program. Arrange to show, twice, commercial A in one city and commercial B in the other. After the program has been aired, conduct random telephone surveys in the two cities to measure viewer reactions. Specifically, in each city, measure preference for GFPC's frozen pizzas among a sample of 50 consumers who saw the program. Compare the mean consumer preference for GFPC's frozen pizzas for the samples in the two cities to assess the impacts of commercial A and commercial B.

What similarities and differences do you see between approaches 1 and 2? Both are experimental in that they involve manipulating a causal variable (namely, the type of test commercial shown) and measuring its impact on an effect variable (namely, consumer preference for GFPC's frozen pizzas). A major difference between the two approaches, however, is the degree of control available during the manipulation and measurement process. Approach 1 clearly offers better control than approach 2 with respect to extraneous factors capable of influencing consumer preferences toward GFPC's frozen pizzas. Approach 1 is a form of laboratory experiment, while approach 2 is a form of field experiment.

Definition. A ***laboratory experiment*** is a research study conducted in a contrived setting in which the effect of all, or nearly all, possible influential independent variables not pertinent to the immediate problem is kept to a minimum.

Definition. A ***field experiment*** is a research study conducted in a natural setting in which one or more independent variables are manipulated by the experimenter under as carefully controlled conditions as the situation will permit.[7]

---

[7] The two definitions are adapted from those given in Kerlinger, *Foundations of Behavioral Research,* pp. 398, 401.

The key procedural difference between the two types of experiments, in terms of the research setting and hence the degree of control over extraneous factors, has an important bearing on the validity of the experimental results. The validity of experimental results is usually evaluated on two dimensions: external validity and internal validity.[8] Campbell and Stanley describe these two forms of validity as follows:

> *Internal validity* is the basic minimum without which any experiment is uninterpretable: Did in fact the experimental treatments make a difference in this specific experimental instance? *External validity* asks the question of *generalizability:* To what populations, settings, treatment [independent] variables, and measurement [dependent] variables can this effect be generalized?[9]

Internal Validity *The extent to which results observed in an experiment are solely due to the experimental manipulation.*

External Validity *The extent to which results observed in an experiment are likely to hold beyond the experimental setting.*

Internal validity is the extent to which observed results are solely due to the experimental manipulation; ***external validity*** is the extent to which observed results are likely to hold beyond the experimental setting. An ideal experiment is one whose results will have high internal as well as external validity. Unfortunately, there is usually a trade-off between these two forms of validity. That is, control of all extraneous factors—holding constant other causal variables, or at least estimating and filtering out the influence of such variables on the effect variable—is typically only possible in a contrived setting, which may cast doubt on whether the results apply to realistic settings. Thus laboratory experiments generally have an advantage over field experiments in terms of internal validity but not external validity. The opposite is true for field experiments.

Let us reexamine the two alternative experimental approaches described in the GFPC scenario in light of our discussion of laboratory versus field experiments and internal versus external validity. Suppose *both* approaches were used and the following results were observed:

- Approach 1 (laboratory experiment). In the group exposed to commercial A, 60% preferred GFPC's frozen pizzas over competing brands. Consumer preference for GFPC's frozen pizzas in the group exposed to commercial B was 40%.
- Approach 2 (field experiment). Consumer preference for GFPC's frozen pizzas was 25% among those exposed to commercial A and 30% among those exposed to commercial B.

---

[8] Although these forms are the two most commonly used forms of validity in experimental research, they are by no means the only forms; for definitions and a comparative discussion of other forms of validity, see Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design & Analysis Issues for Field Settings* (Chicago: Rand McNally, 1979), pp. 80–91.

[9] Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1963), p. 5.

The results from the two experiments are obviously at odds with each other. Can such inconsistent results occur in this scenario? They certainly can, since neither experiment can claim high internal *and* external validity.

The 20-point difference in preference between the two groups in the laboratory experiment can be attributed confidently to the two different test commercials because virtually everything else remained strictly controlled. The results therefore have high internal validity. There is some question, however, about whether commercial A would be so much more effective than commercial B in a real-life setting, given the artificial environment in the laboratory experiment. For instance, consumers watching television under normal circumstances may not pay as keen attention to the program and the commercials in it, and, as a result, consumer preferences may not differ to as great an extent as in a laboratory experiment.[10] Thus the 20-point difference in consumer preference cannot be generalized confidently.

The results of the field experiment, giving commercial B a slight edge over commercial A, may suffer from a different type of problem. While the experimental setting itself was quite realistic, uncontrolled factors (in addition to the test commercial) may have influenced consumer preferences in different ways in the two cities. For instance, competing brands of frozen pizza may have been heavily promoted on the day of the experiment in the city in which commercial A was shown but not in the city in which commercial B was shown. If so, the finding that commercial B is somewhat more effective than commercial A will lack internal validity and cannot be fully trusted.

Which of the two approaches should the advertising agency use? Or, more generally, are laboratory experiments likely to be better than field experiments, or vice versa, in terms of overall usefulness of the results? There are no clear-cut answers. As we will see in the next section, a variety of situational factors play a role in the choice of an appropriate experimental approach. Nevertheless, the preceding discussion and illustrations have an important message to all researchers using marketing experiments: No matter how much control and realism you *think* you have incorporated into your experiment (be it a laboratory or a field experiment), unanticipated or uncontrolled conditions may influence the results, and therefore you should temper your inferences accordingly when such conditions do crop up.

---

[10] Indeed, several research studies have shown that laboratory experiments have a tendency to overestimate the effect of experimental manipulations. For an example of such a study involving price manipulations in a laboratory setting, see John R. Nevin, "Laboratory Experiments for Estimating Consumer Demand: A Validation Study," *Journal of Marketing Research,* 11 (August 1974), pp. 261–268.

DECIDING ON WHICH TYPE OF EXPERIMENT TO USE

So far we have considered only validity issues relating to laboratory versus field experiments. Other practical considerations relevant in choosing between the two types of experiments include time, cost, exposure to competition, and nature of the manipulation. These issues are addressed in the following subsections.

Time

As you might suspect, conducting an experiment in a real-life setting will generally be more time-consuming than conducting it in a contrived setting. A field experiment will usually require additional time for such activities as identifying appropriate field locations and introducing the experimental manipulations into those locations. For instance, consider approach 2 in the GFPC scenario. Selecting two similar test cities, identifying an appropriate television program to be aired simultaneously in both cities, purchasing commercial airtime on that program, and so on, are quite likely to be time-consuming tasks.

Cost

With cost, again, laboratory experiments generally have an advantage over field experiments. The larger scale on which manipulations have to be introduced and results monitored in field experiments can sharply increase their cost. For instance, in the GFPC scenario, buying commercial airtime on television programs in the two test cities is likely to make approach 2 significantly more expensive than approach 1. Given a situation requiring experimental research, a laboratory experiment can usually be conducted for a fraction of the cost of running a field experiment.

Exposure to Competition

A potential drawback of any marketing field experiment is the danger of exposing it to competitors. Exposure may invite competitors to monitor the experiment and gain valuable insights virtually for free. It may also prompt competitors to change the experimental environment (e.g., by drastically changing their own promotional or pricing levels) and hence invalidate the experimental results. Clearly, competitive monitoring and interference in an experiment are highly undesirable from the experimenter's perspective.

Test Marketing *A form of field experiment for assessing a market's reactions to a new product and its associated marketing mix.*

The threat of competitive interference is especially great in ***test marketing***, which is a form of field experiment for assessing the market's reactions to a new product and its associated marketing mix. Test marketing is a formal step prior to full-scale introduction in the new-product-planning processes of many consumer goods firms.[11] But the risk of competitive interference does not necessarily mean that laboratory experiments must be used instead of field experiments for gaining insight into the potential market impact of new products. Consumer response to new products, especially major innovations (e.g., videocassette recorders and videodisc players), simply cannot be gauged accurately in a laboratory environment. Because of consumers' lack of familiarity with new products, their reactions to conceptual descriptions or even prototypes of the product in a laboratory setting may be no better than wild guesses, which are unlikely to have external validity.[12] Furthermore, laboratory experiments cannot adequately duplicate important phenomena like diffusion of advertising messages through various groups of consumers and the effect of manufacturer's advertising on the inventory decisions of wholesalers and retailers, which, in turn, can influence sales.[13] The point of this discussion is that test marketing, despite the risk of exposure to competition, may still be the most meaningful approach in certain situations. A key to avoiding drawing erroneous inferences from test market results is to take into account any unusual competitive activity when one is interpreting the results.[14]

Accurate assessment and interpretation of competitive activity during a test-marketing operation may not always be easy, however. The following example illustrates this point.

Example. When Procter and Gamble test-marketed its Wondra brand hand-and-body lotion, a leading competitor, Cheseborough Ponds, ran a buy-one-get-one-free promotion for its Vaseline Intensive Care Lotion in the test cities. Needless to say, the sales performance of Wondra was quite poor. Nevertheless, Procter and Gamble apparently felt that Ponds could not afford to run such a promotion on a national basis. It therefore went ahead with full-scale introduction of Wondra, despite its disappointing test market performance. Procter and Gamble's assessment of Ponds' promotional capability turned out to be inaccurate; Ponds ran a buy-one-get-one-free promotion wherever Wondra was introduced. Consumers stocked up on Vaseline Intensive Care Lotion, which seriously hurt Wondra's chances of establishing a strong foothold in the hand-and-body-lotion market. Eight years after national introduction of Wondra, the brand still had only a 4% market share, while Vaseline Intensive Care had 22%.[15]

---

[11] For a succinct discussion of test marketing, see Philip Kotler, *Marketing Management: Analysis, Planning and Control* (Englewood Cliffs, N.J.: Prentice-Hall, 1980), pp. 336–339; another excellent reference concerning test marketing is N. D. Cadbury, "When, Where, and How to Test Market," *Harvard Business Review*, May/June 1975, pp. 96–105.

[12] A thought-provoking article that makes this point forcefully, with several excellent examples, is "Marketing Researchers: Investigate Why New Products Succeed, Not Why They Fail," *Marketing News*, September 18, 1981, Section 1, p. 13; see also Edward M. Tauber, "Forecasting Sales Prior to Test Market," *Journal of Marketing*, 41 (January 1977), p. 82.

[13] Paul W. Farris and David J. Reibstein, "Overcontrol in Advertising Experiments," *Journal of Advertising Research*, 24 (June/July 1984), pp. 37–42.

[14] Several useful precautions to take when one is conducting test-marketing experiments are discussed in Thomas D. DuPont, "How to Avoid Test-Marketing 'Traps'," *Marketing News*, May 13, 1983, Section 2, pp. 16–18.

The point of this example is that proper interpretation of test market results requires good knowledge of not only how competitors react during the test-marketing operation but also what they are capable of doing in the future.

Nature of the Manipulation

The type of independent variable used and how it is to be manipulated also have a bearing on which experimental approach will be most appropriate. Laboratory experiments will be more appropriate than field experiments—especially given the time, cost, and secrecy advantages of the former—*as long as the manipulations are likely to be meaningful in a contrived setting*. For example, laboratory experiments may be quite appropriate for assessing consumer reactions to different types of package designs for a soft drink, different levels of flavoring added to an ice cream, and so on.

Field experiments are better when the manipulation is likely to be meaningless or difficult to implement in a laboratory setting. We have already seen that the potential market impact of major innovations cannot be adequately assessed through laboratory experiments because of the difficulty in evoking meaningful reactions from respondents. Lodish and Pekelman offer additional examples of situations where the nature of the manipulations will make laboratory experiments inappropriate:

> Though laboratory experiments may be useful in some situations, certain types of changes in the system can be investigated only by means of field experimentation. These changes are very difficult to duplicate realistically in the laboratory. Common examples are changes in salesmen's compensation, in the organization of the salesforce, in promotional methods used by salesmen, in the structure of the distribution system, and in pricing policies or credit terms. The most common experimental approach is to implement the change in some subset of the present territories, and then to compare the performance of these territories with that of the ones not receiving the treatment.[16]

To summarize our discussion of the pros and cons of laboratory versus field experiments: Laboratory experiments are strong in terms of internal validity (ability to determine the sole impact of a causal variable), time, cost, and negligible risk of exposure to competition. Field experiments are strong in terms of external validity (generalizability of inferences) and meaningfulness of the manipulations. To capitalize on the strengths of both types of experiments, and to circumvent the difficulty of making a strictly either-or decision, researchers have developed hybrid experimental approaches. These hybrid approaches are essentially laboratory experiments that imitate real market conditions as much as possible. They are therefore often called ***simulated test markets***. Several commercial marketing research firms have facilities especially designed for conducting experiments under conditions closely resembling actual market conditions. One such facility, called the Litmus℠ Simulated Test Market (LSTM), is offered by Yankelovich Clancy Shulman, a well-known marketing research firm. The LSTM is part of a market-response research and forecasting approach called the Litmus™ system. A description of the LSTM and how it is used is summarized in Research in Use 8.1.

---

[15] "How to Improve Your Chances for Test-Market Success," *Marketing News*, January 6, 1984, Section 1, pp. 12–13.

[16] Leonard Lodish and Dov Pekelman, "Increasing Precision of Marketing Sales Areas," *Journal of Marketing Research,* 15 (August 1978), p. 449.

Simulated Test Market
A hybrid experimental approach involving a laboratory setting that imitates real market conditions as much as possible.

Simulated test markets are by no means completely devoid of artificiality. However, a few studies comparing results obtained from simulated and actual market experiments have shown a moderate degree of consistency between the two sets of results.[17] Experiments conducted in simulated test markets are likely to be more frequently used in the future because they are becoming increasingly sophisticated and also because they bring together the complementary strengths of laboratory and field experiments.[18] In the GFPC scenario, for instance, testing the relative impacts of commercials A and B in a simulated market (such as the one described in Research In Use 8.1) may be better than either of the two proposed approaches. The simulated test market results will yield estimates of the immediate as well as the longer-term impacts of the commercials on *sales* (rather than on *preference* alone). Thus the information

---

[17] See, for example, Glen I. Urban and Gerald M. Katz, "Pre-Test-Market Models: Validation and Managerial Implications," *Journal of Marketing Research,* 20 (August 1983), pp.. 221–234; Andre Gabor, Clive W.J. Granger, and Anthony P. Sowter, "Real and Hypothetical Shop Situations in Market Research," *Journal of Marketing Research,* 7 (August 1970), pp. 355–359.

[18] Given the ability of laboratory experiments to detect even minute differences in the impact of various manipulation levels, some authors have suggested the following alternative approach for taking advantage of the complementary strengths of laboratory and field experiments: Use laboratory experiments as a *prelude* to field experiments. Specifically, when the effects of several independent variables have to be studied, the ones that are not likely to make much difference can be identified through laboratory experiments and eliminated from further consideration. The effects of the remaining independent variable(s) can be studied in a field experiment, which will be much less complex and cumbersome than it would have been had all the original independent variables been manipulated in a field setting. See Alan G. Sawyer, Parker M. Worthing, and Paul E. Sendak, "The Role of Laboratory Experiments to Test Marketing Strategies," *Journal of Marketing,* 43 (Summer 1979), pp. 60–67.

Research in Use 8.1 *Yankelovich Clancy Shulman*

Litmus™ Simulated Test Market

In a simulated test study, we test consumers with the actual product or service in an *actual purchase situation*, to observe their actual trial, repeat, and usage behavior, and then forecast the results in the real world. (This is not a coupon redemption study, or self-reports on a 5-point scale. It's *real behavior*.)

Our Simulated Test Marketing puts carefully screened consumers into a special test store or service situation. The store is stocked with competitors' products as well as yours, based on location surveys. These customers use their own money to buy (true simulation), after being exposed to your ad campaign and to your competitors'. They are questioned on attitudes before going into the store, and on reactions (a) after the ad-campaign exposure, (b) after the store experience, (c) after home usage of whatever products or services they buy, and (d) at continued repurchase stages.

This response data yields forecasts of trial, first and second repeat, usage rate, and other measures for use in the Litmus model.

The market-response testing phase here also provides information on consumer diagnostics, user and nonuser demographic profiles, and the sources of your volume.

Because in Simulated Test Marketing we are observing actual consumer behavior we can use the Litmus model to develop for you the detailed price/marketing-plan combination that will give you optimal profits as well as sales.

THE TESTING SEQUENCE

Step #1 Prerecruitment.

- Appropriate sample
- Telephone
- Security screen

Step #2 Background: habits and practices.

- Security screen
- Self-administered questionnaire
- Current behavior
- Usage occasions and frequency
- Diagnostics
- Rating of ideal attributes and benefits
- Demographics

Step #3 Exposure to real advertising in a competitive context.

(Continued)

Research in Use 8.1 *Continued*

Step #4 Simulated store purchase reflective of BDI/CDI issues.

- Shelf set and price matched to each market
- Consumers make real money purchase

Step #5 Post-purchase inquiry.

- Perceived ad message
- Reaction to price and packaging
- Anticipated purchase and usage
- Rating test with competitive brand(s) on delivery of attributes and benefits
- Likes/dislikes

Step #6 Respondents take product home (no reference to follow-up interviewing).

Step #7 Post-usage evaluation.

- Telephone interview
- Reaction to product
- Degree (and reasons) for satisfaction/dissatisfaction
- Usage pattern
- Rating of brand on attributes
- Attitudinal and behavioral repeat and usage measurements

Step #8 Extended usage evaluation(s).

- As needed
- Repeat Step #7

Source: Excerpted from "Market Response Research and the Litmus™ System," a brochure prepared by Yankelovich Clancy Shulman, 8 Wright Street, Westport, Conn. 06880, p. 9 and pp. 13–14. Used by permission.

generated is likely to be richer and managerially more useful than that generated in a traditional laboratory experiment.

A recent development that is enabling marketers of packaged goods to conduct sophisticated field experiments is the availability of scanner data—the data on product sales through supermarkets equipped with electronic scanners at the checkouts. Several marketing research firms have established panels of scanner-equipped stores spread over areas that are representative of the major markets in the country. The data from these stores are transmitted electronically to central computers for analysis and interpretation. Retailers and manufacturers can systematically manipulate marketing stimuli (e.g., price, promotion, shelf space allocation) and ascertain their impact on product sales by analyzing scanner data from the markets participating in the experiment. The A. C.

Nielsen Company, a leader in the area of scanner-based research, offers marketers a variety of services through its information system called ScanTrack.[19]

We have so far seen specific illustrations of biases (primarily in the context of the GFPC scenario) that can lower the validity of experimental results. In the next two sections of the chapter we will examine certain standard threats to the internal and the external validity of experiments in general. As you examine those threats, keep in mind that not all of them may be present in every experiment. The type of experiment (laboratory or field) and the type of units from which data are gathered (whether they are human respondents or nonhuman entities such as stores) will have a bearing on which threats are likely to be serious. Furthermore, as we will see in the next chapter, some of the threats, even if they cannot be completely removed from an experimental setting, can be neutralized or accounted for by choosing an appropriate experimental design.

THREATS TO INTERNAL VALIDITY

Internal validity, as we saw earlier, refers to the extent to which one can be confident that the manipulated independent variable is solely responsible for observed changes in the dependent variable. Therefore the presence of any condition or occurrence (other than the independent-variable manipulation) that can offer a competing explanation for the experimental results is a threat to internal validity. We will discuss the following effects (i.e., conditions or occurrences) that can lower the internal validity of experiments: history, maturation, pretesting, instrument variation, selection, and mortality.[20]

History Effect

History Effect *A bias that occurs because of specific external events or occurrences during an experiment (e.g., changes in competitors' marketing mixes) that are likely to affect the dependent variable.*

History effect refers to *specific* external events or occurrences during an experiment that are likely to affect the dependent variable. The history effect is not a frequent problem in laboratory experiments. Because of the controlled environments in which such experiments are conducted, the researcher is in a good position to prevent, or at least monitor and account for, any extraneous event or occurrence other than the planned manipulation. The history effect is potentially a more serious problem in field experiments. Unusual shifts in competitors' marketing mix variables when a field experiment is in progress are the most likely form of the history effect.

---

[19] For additional information about ScanTrack and the services it offers, see "Nielsen Scanning-Based Information Systems—A New Generation of Research," *The Nielsen Researcher,* Vol. 1, No. 1, Fall 1986, pp. 2–9, published by Nielsen Marketing Research, Nielsen Plaza, Northbrook, Ill. 60062.

[20] This list is not an exhaustive list of threats to internal validity. Rather, it represents threats that are most likely to be present in experimental designs used in marketing, particularly those designs to be discussed in the next chapter. The book by Campbell and Stanley, *Experimental and Quasi-Experimental Designs,* and the one by Cook and Campbell, *Quasi-Experimentation,* provide a more comprehensive coverage of threats to internal validity.

Suppose the producer of Dole canned fruits wishes to study the impact of a special store display on the brand's sales. A representative group of stores in a test area is selected, and sales of Dole canned fruits are monitored in these stores for a one-month period. The special display is then introduced, and sales are monitored for another month. Ideally, the impact of the special display should be revealed by the difference between the second and first month's sales of Dole canned fruits in the test stores. But what happens if the prices or promotional strategy for Del Monte canned fruits change significantly during the two-month experimental period? Or if Del Monte experiences distribution problems during that period and stocks out in several of the test stores? Occurrences like these are history effects and may be responsible, at least partly, for any difference between sales of Dole canned fruits with and without the special display. Consequently, the experimental findings cannot be taken at face value.

The history effect is not limited to changes in competitors' marketing mixes or strategies. It can also occur because of major changes in external factors like the weather or general economic conditions in the area where a field experiment is being conducted. Examples are a prolonged heat wave in an area where the effect of a price reduction on sales of a brand of soft drink is being tested, and a series of plant closings and layoffs in an industrialized area where the effect of a promotional campaign on sales of a brand of microwave ovens is being tested.

The type of units (human or nonhuman) from which data are gathered during an experiment does not really have a bearing on the extent of the history effect. The history effect stems from the experimental environment rather than the experimental units. To illustrate, compare the field experiment in the GFPC scenario to test the impact of commercials A and B on consumer preferences with the field experiment to test the impact of the special store display on sales of Dole canned fruits. In the GFPC experiment preference data are obtained from consumers, while in the Dole experiment brand sales data are gathered from stores. Despite this difference in type of units, however, any history effect—for example, a change in a leading competitor's advertising campaign—can be expected to affect the dependent variable in both experiments (consumer preference in the former and brand sales in the latter).

Maturation Effect *A bias that occurs because of the impact on the dependent variable of physiological or physical changes in units participating in an experiment.*

Maturation Effect

Unlike the history effect, which stems from the external experimental environment, the ***maturation effect*** stems from the experimental units. The term *maturation* refers to physiological or physical changes in the units that occur with the passage of time. Any impact such changes may have on the dependent variable being measured constitutes a maturation effect.

Consider a laboratory experiment to ascertain the impact of a new Toyota commercial on consumer opinions about Toyota cars. A representative group of consumers is brought to a laboratory setting, and their current opinions about Toyota cars are measured through an appropriate questionnaire. They are then asked to watch an hour-long television program in which the test commercial is inserted. Next, their opinions about Toyota cars are measured once again. Could any physiological changes have occurred in the respondents during this experiment? Yes, since some of them may have started feeling hungry toward the end of the experiment, others may have become tired, and so on. If such changes prompted the respondents to "just get the whole thing over with," their opinions as measured through the second questionnaire would be inaccurate and the maturation effect would be present. In other words, the difference between the pretest and posttest measures of opinions cannot be completely attributed to the test commercial; part of the difference may be due to maturation.

Maturation is not limited to changes in the hunger and fatigue levels of respondents. It also involves changes that naturally occur with the passage of time, such as changes in respondents' ages and their ability to react quickly to stimuli. However, such changes occur so gradually that they are unlikely to pose a problem in marketing experiments, which are invariably of relatively short duration. In contrast, physical conditions like the respondents' fatigue and hunger levels can change rapidly, especially when respondents are placed in an unusual environment. This point highlights the possibility of the maturation effect's occurring in the contrived settings that characterize laboratory experiments. The longer the duration of a laboratory experiment and the more unusual the respondents' experiences during the experiment, the greater is the risk of reduced internal validity owing to some form of maturation effect.

The maturation effect can also occur when the units participating in an experiment are nonhuman (e.g., stores, as in the field experiment for sales of Dole canned fruits). For instance, the layout or other physical features of test stores can change over time, which in turn can influence sales. Or the overall sales in the test stores may be following a certain general trend (upward or downward), which, if not recognized and controlled for during an experiment, may be mistakenly viewed as an effect on the experimental manipulation.

Pretesting Effect

The pretesting effect is a problem unique to experiments involving human participants in which the same respondents are measured more

***Pretesting Effect** A bias in experimental studies that occurs when responses given during a later measurement are influenced by those given during a previous measurement, regardless of what happens between the measurements.*

than once. Specifically, the ***pretesting effect*** is said to occur when responses given during a later measurement are influenced by those given during a previous measurement, regardless of what happens between the measurements. When the pretesting effect occurs, the difference between the pretest and posttest measures of the dependent variable will not accurately reflect the impact of the experimental manipulation, thereby lowering internal validity.

Recall the laboratory experiment to measure the impact of a new Toyota television commercial. In this experiment, suppose consumers' opinions before and after exposure to the commercial are measured through a series of 10-point rating scales relating to various attributes of Toyota (rating scales and their development are discussed in Chapter 11). While responding to the second measurement, the respondents may remember how they rated Toyota on each scale during the first measurement. Furthermore, some of these respondents may provide the same ratings they gave in the previous measurement, perhaps through a desire to appear consistent. If so, the difference between the two measurements will reveal nothing about the true impact the Toyota commercial may have had on these respondents. Alternatively, some respondents may provide a different set of ratings during the second measurement, not because their opinions about Toyota cars had changed but because they did not want to give the same set of ratings twice. In either case the result is a lowering of internal validity: The difference, or lack thereof, between the pretest and posttest opinions about Toyota cars cannot be attributed solely to the Toyota commercial.

The pretesting effect stems from the tendency (for whatever underlying reasons) to remain consistent or to reveal differences between the pretest and posttest measurements. Since this tendency is purely a human phenomenon, the pretesting effect is rarely a threat to the internal validity of experiments in which measurements are obtained from nonhuman units. Thus, for example, the pretesting effect is unlikely in the experiment for measuring the impact of a special store display on sales of Dole canned fruits, even though this experiment involves both a pretest and a posttest measurement.

Instrument Variation Effect

***Instrument Variation Effect** A bias in experimental studies that relates to differences between pretest and posttest measurements owing to changes in the instruments (questionnaires) and/or procedures used to measure the dependent variable.*

Instrument variation effect is a bias that relates to differences between pretest and posttest measurements owing to changes in the instruments (questionnaires) and/or procedures used to measure the dependent variable. It is a potential internal-validity threat only in experiments involving more than one measurement of the same dependent variable. However, it is unlikely to be a problem when the units from which measurements are obtained are nonhuman (e.g., stores). That is, dependent variables on which data are obtained from nonhuman units are typically straightforward (e.g., sales, inventory turnover, brand share), with little room for variation in how they are measured. Of course, if the definitions of these variables or the ways in which they are measured change during the experiment, the instrument variation effect will be a problem. But to the extent that such changes can be prevented by the researcher in measuring nonhuman units, the threat of the instrument variation effect is unlikely to be serious.

The instrument variation effect is a much more serious threat to the internal validity of experiments involving the measurement of human respondents. Dependent variables in such experiments are invariably nebulous constructs like attitudes, opinions, preferences, or purchase intentions. Measuring such variables typically requires a questionnaire and calls for some form of interaction between respondents and the interviewer. Any changes in the questionnaire or the interviewer between pre- and postmeasurements can, in and of itself, produce a difference between the results obtained.[21]

Ideally, the experimenter should use identical data collection procedures for pre- and postmeasurements to avoid the problem of instrument variation effect. Unfortunately, in some situations changes in data collection procedures between measurements may be inevitable. For example, changes in wording may be necessary in the postmeasurement questionnaire because respondents had difficulty understanding some questions in the premeasurement questionnaire. (A good way to minimize the need for such changes is to thoroughly pretest the questionnaire before finalizing it. We will discuss the topic of questionnaire pretesting—not to be confused with the pretesting effect—in Chapter 10.) Such changes will obviously increase the chances of obtaining an instrument variation effect. Sometimes, a different interviewer might be needed to conduct the posttest measurement because the original interviewer quit or was incompetent. For instance, consider an in-home product-testing experiment, in which respondents are initially interviewed in their homes concerning their views about a product; then samples of the product are left with respondents for their use over a period of time; at the end of this period the respondents are reinterviewed to ascertain whether and to what extent their opinions have changed. Instrument variation effect can be a problem here if the person conducting the second interview is different from the person who conducted the initial interview.

In summary, an experimenter should use the same measurement procedures if at all possible. If some changes are inevitable, the experimenter must be especially alert to the possibility of instrument variation effect and take it into account in interpreting the results.

---

[21] As we will see in Chapter 10, even slight changes in a questionnaire can lead to big differences in responses. For a discussion of the extent to which responses can be influenced by the characteristics of the interviewer, see M. Venkatesan, "Laboratory Experiments in Marketing: The Experimenter Effect," *Journal of Marketing Research,* 4 (May 1967), pp. 142–146.

Selection Effect

Selection Effect *A bias that occurs when multiple groups participating in an experiment differ on characteristics that have a bearing on the dependent variable.*

Selection effect is a potential problem in experiments involving more than one group of units. This bias is present when multiple groups participating in an experiment differ on characteristics that have a bearing on the dependent variable. For instance, the laboratory as well as the field experiment approaches proposed in the GFPC scenario to test the relative effectiveness of commercials A and B involved two groups of respondents. Suppose the group exposed to commercial A was dominated by heavy users of frozen dinners, while the group exposed to commercial B was dominated by consumers who typically ate only home-cooked meals. Also, suppose 40% of those who saw commercial A indicated a preference for GFPC's frozen pizzas while only 5% of those who saw commercial B did so. These results cannot necessarily be interpreted as showing commercial A to be more effective than commercial B. All or part of the observed difference in preferences may be due to the critical difference between the general food preferences of the two groups rather than to the experimental manipulation (i.e., the difference between the two commercials). The internal validity of the experiment under these circumstances is likely to be quite low because of the presence of selection effect.

Selection effect is not limited to only those experiments involving human participants. To illustrate, recall the example we discussed earlier concerning sales of the Duran Duran album *Rio* in various sections of Dallas. The implied inference in this example was that the availability of MTV was responsible for the sales success of the album. We have already seen that this inference is likely to be erroneous owing to the lack of control of extraneous factors. But let us reexamine this situation from a somewhat different perspective to illustrate the concept of selection effect. This experiment used two groups of nonhuman units: Sections of Dallas that had MTV made up one group, and those that did not have MTV made up the other group. The sales results showed that the album was a big success in the former group but not in the latter. However, this difference cannot be attributed with assurance to the MTV–no-MTV distinction because the two areas could have differed significantly on characteristics like socioeconomic status and interests of residents. Stated differently, no attempt was made to ensure that the two areas were equivalent on attributes (other than availability of MTV) likely to affect the album's sale. The possibility of selection effect is a primary reason why the inference implied in this situation cannot be fully trusted.

Mortality Effect

Mortality Effect *A bias that occurs when certain participating units drop out of an experiment, and, as a result, the set of units completing the experiment significantly differs from the original set of units.*

Mortality effect occurs when certain participating units drop out of an experiment, and, as a result, the set of units completing the experiment significantly differs from the original set of units. To illustrate the mortality effect, we will use the in-home product-testing experiment mentioned earlier (under "Instrument Variation Effect"). Suppose the product is a new and improved bathtub and tile cleaner, and the initial set of participants consists of 100 users of the current version of the product. In the pretest interview these participants are asked to state their overall opinion of the old version of the bathtub and tile cleaner on a scale of 1 to 10 (the higher the number, the more favorable is the opinion). Assume that the mean rating across all 100 participants turns out to be 6. The participants are then given samples of the new and improved version and asked to use it over a two-month period. At the end of this period, suppose only 70 participants are reinterviewed because the rest simply refuse to be interviewed. Assume that the mean rating across the 70 participants is 8 on the same 10-point scale. The key question now is, Can the two-point improvement in the mean opinion rating be attributed to the new and improved version?

The answer to this question critically depends on the *composition* of the 30 dropouts vis-à-vis the composition of the original set of participants. For instance, if the 30 that dropped out were people with negative reactions to the product, the postmeasured mean rating of 8 could be inflated. In other words, the difference in mean opinion ratings is not a true reflection of the effectiveness of the new and improved version because the composition of participants changed critically between measurements. Thus mortality effect is present.

Since 30 participants dropped out, why not simply recalculate the initial mean opinion rating over just the 70 participants that completed the experiment and compare it with the mean rating from the postinterviews? Will this procedure solve the mortality effect problem? Unfortunately it will not; while this procedure will ensure that the pretest and posttest mean ratings are based on the same group of participants, inferences stemming from such a comparison cannot be generalized to the original group or to the larger body of product users the original group represents. In short, the inference will lack external validity. Thus while the mortality effect is typically viewed as a threat to internal validity, it can become a threat to external validity if the experimenter attempts to circumvent the problem by making inferences on the basis of data from only those units that completed all phases of the experiment.

Of course, if the 30 people that dropped out were random dropouts—that is, if the characteristics of those dropping out had no systematic connection to the dependent variable—the pre- and postinterview groups would differ in size but not in composition. Hence there would be little or no mortality effect. This last point is important because the mortality effect is not necessarily present whenever a loss of participants occurs; it is a problem only when such a loss results in a group that critically differs from the original group. Therefore an experimenter should collect data on critical participant characteristics (e.g., demographics, extent and frequency of product use) during both the pre- and postinterviews. In the event of loss of participants during an experiment, such data can be valuable in ascertaining whether and in what ways group composition has changed. This information will lower the risk of making erroneous inferences.

Mortality effect can occur in experiments involving human as well as nonhuman participants. For instance, mortality effect can be a threat in the field experiment involving the store display for Dole canned fruits if certain test stores, all belonging to the same supermarket chain, refuse to feature the store display midway through the experiment. The mortality effect is also more likely to be a problem in field experiments than in laboratory experiments because a participant's role in most laboratory experiments is confined to a relatively short duration, with usually no opportunity for dropping out.

THREATS TO EXTERNAL VALIDITY

External validity of experimental results relates to their generalizability. The various internal-validity threats we have discussed also indirectly affect external validity since it would be meaningless to generalize experimental findings that are not even internally valid. In other words, internal validity can be viewed as a necessary but not sufficient condition for external validity. Our discussion in this section will focus on conditions that must be met beyond those needed for internal validity in order for one to conclude that experimental findings have external validity. We will examine biases that stand in the way of generalizing experimental results even when they have high internal validity. Specifically, we will discuss the following three biases: reactive bias, pretest-manipulation interaction bias, and nonrepresentative-sample bias.[22]

Reactive Bias

Reactive Bias *A bias that occurs in experimental studies when participants exhibit abnormal or unusual behavior simply because they are participating in an experiment.*

Reactive bias is the problem of participants' exhibiting abnormal or unusual behavior simply because they are participating in an experiment. This bias is limited to experiments involving human units and is likely to occur only when the units know they are participating in an experiment. Reactive bias can be quite severe in laboratory experiments because the artificial environment and the attention paid to participants by the experimenter are especially conducive to this type of bias. As Kerlinger observes, "Almost any change, any extra attention, any experimental manipulation, or even the absence of manipulation but the knowledge that a study is being done, is enough to cause subjects to change. In short, if we pay attention to people, they respond."[23]

---

[22] As in the case of internal-validity threats discussed in this chapter, these three biases are not the only external-validity threats, although virtually all other external-validity threats are related to these biases. For further details, consult Campbell and Stanley, *Experimental and Quasi-Experimental Designs;* Cook and Campbell, *Quasi-Experimentation.* One point worthy of note concerning the *names* given to the validity threats: External-validity threats are labeled *biases* rather than *effects* merely to distinguish them from internal-validity threats.

Reactive bias typically prompts participants to overreact and to look for minute features or differences in experimental stimuli that they normally would not even notice. This reaction may lead to an internally valid inference (assuming all other internal-validity threats have been controlled) that the experimental manipulation had a significant impact. Such an inference cannot be confidently generalized beyond the experimental setting, however.

Although reactive bias is more likely to occur in laboratory experiments, field experiments involving human subjects are by no means immune to it. In Chapter 4 we discussed consumer panels, which are permanent samples of respondents that are repeatedly measured over time. Many field experiments use data obtained from consumer panels. For example, the impact of a campaign using a 50-cents-off coupon for a brand of detergent can be estimated by comparing purchases of the detergent made by panel members before and after the campaign. A potential problem with this approach is that panel members, *because they know their purchases are being monitored,* may behave differently from normal consumers. In other words, the external validity of inferences based on consumer panel data may be questionable owing to the possibility of reactive bias.

Pretest-Manipulation Interaction Bias

Pretest-Manipulation Interaction Bias *A bias that arises when the premeasurement increases or decreases respondents' sensitivity to the experimental manipulation.*

Pretest-manipulation interaction bias is simply a special form of reactive bias that is unique to experiments involving premeasurement of human participants before exposing them to the experimental manipulation. While reactive bias stems from the influence of the experiment as a whole, pretest-manipulation interaction bias stems from the influence of the premeasurement. Specifically, pretest-manipulation interaction bias will arise when the premeasurement increases or decreases respondents' sensitivity to the experimental manipulation.

To illustrate this bias, and to contrast it with *pretesting effect,* let us reexamine the laboratory experiment to measure the impact of a new Toyota commercial on consumer opinions about Toyota cars. The premeasurement in this experiment will undoubtedly contain questions about Toyota cars. In answering these questions, respondents are quite likely to think a lot about Toyota cars—so much so that they may be entering the next phase of the experiment (i.e., exposure to the TV program containing the Toyota commercial) with a heightened level of sensitivity toward Toyota cars. Alternatively, they may have become so tired of thinking about Toyota cars that they may be entering the exposure phase with a decreased level of sensitivity toward Toyota cars. In either case, because of their altered mental state and sensitivity, their reactions to the Toyota commercial will most likely be different from those of normal consumers who are not premeasured. In particular, the respondents in the experiment may pay abnormally high or abnormally low attention when the Toyota commercial is shown.

---

[23] Kerlinger, *Foundations of Behavioral Research,* p. 345.

The opinion changes (or lack thereof) induced by these abnormal reactions will be reflected by the postmeasurement. Nevertheless, since these reactions still basically stem from the experimental manipulation, the difference between the pre- and postmeasured opinions is attributed to the Toyota commercial *within the experimental context.* In other words, the results will be *internally valid* (assuming that there were no other internal-validity threats). However, they cannot be generalized beyond the experimental setting because the impact of the same commercial on consumers who are not premeasured cannot be expected to be the same.

What is the difference between pretesting effect and pretest-manipulation interaction effect? Very simply, pretesting effect is a *direct* effect that the premeasurement introduces on the postmeasurement, *irrespective of the manipulation.* It is therefore a threat to internal validity. Pretest-manipulation interaction bias, in contrast, is more subtle and occurs because the premeasurement interferes with the mental sensitivity and hence reactions of respondents to the experimental manipulation. This bias ultimately contaminates the postmeasurement, of course. However, since it occurs through an interaction between the premeasurement and the manipulation, rather than because of the premeasurement per se, it is typically considered an effect of the manipulation on the dependent variable. It is therefore regarded as a threat to external validity rather than to internal validity.

Nonrepresentative-Sample Bias

Nonrepresentative-Sample Bias *A bias that occurs in experimental studies when the participating units are not representative of the larger body of units to which the results are to be generalized.*

Nonrepresentative-sample bias occurs when the units participating in an experiment are not representative of the larger body of units to which the experimental results are to be generalized. This bias stems from improper or inadequate recruiting of units and is a sampling problem (problems associated with sampling are presented in Chapters 12 through 15). When the composition of units participating in an experiment critically differs from that of the total collection of units, the experimental results will lack external validity, no matter how internally valid they are.

Nonrepresentative-sample bias can be a potential threat to external validity in all types of experiments, laboratory as well as field, and those involving human and nonhuman participants. It is particularly likely to be present in laboratory experiments and experiments involving consumer panels. There is always the risk that participants willing to expend the effort to take part in a laboratory experiment, or to respond to the repeated measurements to which panel members are subjected, may differ in significant ways from the population at large.[24]

CONCLUDING REMARKS ON VALIDITY THREATS

The preceding two sections covered a number of problems that can lower the internal or external validity of experimental results. These problems, though not exhaustive, are the ones that underlie the practical constraints most frequently encountered in conducting experimental research in marketing.[25] An additional complication concerning these problems is that there are often trade-offs between them; that is, minimizing one problem may aggravate another. For example, an attempt to reduce the pretesting effect by using different questionnaires (i.e., questionnaires with the same content but different formats) for the pre- and postmeasurements is likely to lead to the instrument variation effect. An attempt to circumvent the mortality effect by analyzing data from only those units completing all phases of an experiment may result in nonrepresentative-sample bias. The point is that an experimenter, in striving to avoid certain threats to validity, should not lose sight of the possible increased risk of other threats.

If the number and the variety of potential problems give you the hopeless feeling that experimental research in marketing is useless, do not despair. First, not all problems are likely to be severe in every marketing experiment. As was discussed earlier, the *seriousness* of problems likely to be encountered in a situation depends on the nature of the experiment (laboratory or field) and the nature of the participating units (human or nonhuman). Table 8.1 provides a summary of the circumstances in which each validity threat is likely to be serious.

Second, although a researcher may not be able to eliminate certain threats in a situation, he or she may still be able to account for them by filtering out their influences from the experimental results. In fact, the primary focus of the next chapter is on experimental designs that adequately consider extraneous influences on the dependent-variable measure in assessing the true impact of the causal variable(s).

Third, recall from our discussion at the beginning of the chapter that to conclusively establish causality between variables is virtually impossible. Therefore the realistic role of experimentation in marketing is not so much to *prove* causality as to increase one's confidence in making causal inferences. Viewed from this perspective, carefully conducted experimental research, even if less-than-perfect, can provide more mean-

---

[24] Urban and Katz, "Pre-Test-Market Models," and Gabor, Granger, and Sowter, "Real and Hypothetical Shop Situations," offer several examples illustrating nonrepresentative-sample bias and reactive bias associated with consumer panels.

[25] For a clear and concise review of practical limitations of marketing experiments, see Cox and Enis, *Experimentation for Marketing Decisions,* pp. 108–109.

Table 8.1 *Circumstances Under Which Various Validity Threats May Be Serious*

Type of Threat	Threat May Be Serious for Experiment	 Participants
History effect	Field experiment	Human or nonhuman units
Maturation effect	Laboratory or field experiment	Human or nonhuman units
Pretesting effect	Laboratory or field experiment	Human units
Instrument variation effect	Laboratory or field experiment	Human units
Selection effect	Laboratory or field experiment	Human or nonhuman units
Mortality effect	Field experiment	Human or nonhuman units
Reactive bias	Laboratory experiment	Human units
Pretest-manipulation interaction bias	Laboratory or field experiment	Human units
Nonrepresentative-sample bias	Laboratory or field experiment	Human or nonhuman units

ingful and accurate causal insights than descriptive research can. Stated differently, experimental research in marketing can serve a useful purpose as long as the experimenter recognizes the validity threats that may arise, takes steps to eliminate or control as many of them as possible, and interprets the findings with caution.

SUMMARY

Experimentation is research conducted under controlled conditions to shed light on causal associations between variables. However, the distinction between experimental research and descriptive research is more a matter of degree than of kind. In fact, both types of research typically involve collecting and analyzing similar kinds of data to understand relationships between variables.[26] The distinguishing feature of experimental research is that it can greatly increase one's confidence in making causal inferences.

---

[26] Hans Zeisel, in *Say It with Figures* (New York: Harper & Row, 1968), p. 108, demonstrates with examples that *tabulated results* from both types of research will be virtually indistinguishable.

Three conditions must be satisfied in order to infer that one variable is the cause of another: (1) temporal ordering of variables—that is, the variables (or changes in them) must occur in a time sequence consistent with the causal inference; (2) evidence of association—that is, data on the two variables must show a definite relationship; and (3) control of other causal factors—that is, all extraneous influences on the two variables and their relationships must be adequately accounted for.

Two approaches are available for conducting experimental research: laboratory experiments and field experiments. A laboratory experiment is conducted in an artificial setting under tightly controlled conditions. It therefore generally offers better internal validity (i.e., assurance that the results obtained are solely due to the experimental manipulation) than a field experiment, which is conducted in a natural setting with as much control as possible. A key drawback of laboratory experiments is that the results obtained may not be generalizable beyond the experimental setting, thereby lowering external validity.

The choice between laboratory and field experiments is not always easy because of the trade-off involved between internal and external validity. Furthermore, issues such as time, cost, risk of exposure to competition, and nature of the manipulation must also be considered in choosing an appropriate experimental approach. In some instances a researcher may be able to use a hybrid approach that pools the complementary strengths of laboratory and field experiments. Several commercial marketing research firms have facilities, called simulated test markets, especially designed for conducting hybrid experiments. An experiment conducted in a simulated test market is a laboratory experiment in which the environment and the procedures used imitate real-life market features as much as possible.

There are six standard threats that can jeopardize the internal validity of experiments: history, maturation, pretesting, instrument variation, selection, and mortality effects. These threats also indirectly affect external validity because generalizing results that lack internal validity is meaningless. Direct threats to external validity include reactive bias, pretest-manipulation interaction bias, and nonrepresentative-sample bias.

Not all threats to internal and external validity may be present in every experiment. Whether an experiment is a laboratory or a field experiment, and whether the participants are human or nonhuman units, play a role in determining which threats are serious. Specific experimental designs capable of either eliminating or accounting for the various threats are discussed in the next chapter.

KEY TERMS

Experiment
Laboratory experiment
Field experiment
Internal validity
External validity
Test marketing
Simulated test market
History effect
Maturation effect

Pretesting effect
Instrument variation effect
Selection effect
Mortality effect
Reactive bias
Pretest-manipulation interaction bias
Nonrepresentative-sample bias

QUESTIONS

1. Are descriptive and experimental research totally different? Why or why not? Explain your answer with an example of your own.
2. Toys & Games is a children's toy store located in a shopping mall. Toys & Games ran a major in-store promotion (consisting of special displays and price reductions) during the last quarter (October to December) of each of the past five years. Every year, without fail, its sales in the last quarter were at least 25% higher than sales in any of the previous quarters. Thus the store manager is seriously thinking about expanding the in-store promotion to at least one other quarter each year. What advice would you give the store manager? Discuss your answer from the perspective of the three conditions for inferring causality.
3. Distinguish between laboratory and field experiments, pointing out the relative strengths and weaknesses of each.
4. A food products company has designed a new and improved container for its brand of peanut butter that will reduce waste and maintain freshness over a longer period of time. However, this container will increase the product's price by 30%. The company wants to ascertain whether and to what extent the new and improved container will affect sales of its peanut butter. Propose an appropriate experimental approach that can provide the information desired. Defend your answer.
5. Apart from validity considerations, what factors play a role in the choice between laboratory and field experiments?
6. Briefly describe simulated test markets and their strengths.
7. In what ways does the history effect differ from the maturation effect?
8. "Instrument variation effect will not be a problem as long as the researcher is very careful and consistent." Discuss the extent to which this statement is likely to be true.
9. Can the mortality effect lower *external validity* in any way? Why or why not?
10. Why is the maturation effect more likely to be a problem in laboratory experiments than in field experiments?
11. Refer to the situation described in Question 2, and assume that the research conducted by Toys & Games is an experiment. Is this experiment a laboratory or a field experiment? Does it involve human or nonhuman participants? Go through the list of internal- and ex-

ternal-validity threats we discussed and indicate which are and which are not likely to be serious in this experiment.

12. In what way does reactive bias differ from pretest-manipulation interaction bias?
13. Describe, with an example of your own, the distinction between pretesting effect and pretest-manipulation interaction effect.

CHAPTER 9

Selected Experimental Designs

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between pre-experimental, true experimental, and quasi-experimental designs.*
- *Compare and contrast the following pre-experimental designs: one-group, after-only design; one-group, before-after design; and two-group, ex post facto design.*
- *Explain why random assignment is generally a better procedure than matching for allocating units to various experimental and control groups.*
- *Distinguish between random assignment and random sampling.*
- *Compare and contrast the following true experimental designs: two-group, before-after design; two-group, after-only design; and four-group, six-study design.*
- *Discuss when a two-group, after-only design would be better than a two-group, before-after design.*
- *Illustrate the use of a panel design and explain how this design offers a greater degree of control than pre-experimental designs.*
- *State what is meant by a "statistical" design.*
- *Describe the distinguishing features of the following statistical designs: completely randomized design; randomized-block design; Latin-square design; and factorial design.*

The previous chapter presented an overview of experimental research, with emphasis on the conditions to be met for making causal inferences, the laboratory versus field approaches for conducting marketing experiments, and the various internal- and external-validity threats to experimental findings. In this chapter we will discuss several standard experimental designs that frequently serve as the basis for causal inferences in marketing. Before we examine those designs, a few points concerning them are in order.

First, most of the standard designs can be employed in laboratory or field settings and can involve human or nonhuman participants. Therefore although each design has certain inherent weaknesses, whether and to what extent those weaknesses are likely to be serious may depend on the particular context in which the design is employed. (Recall that the seriousness of several of the validity threats depends on these features; see Table 8.1.) Second, the experimental designs we will look at are not the only ones available.[1] Third, depending on the unique feature of a given situation and the experimenter's ingenuity, the experimenter may be able to create an *ad hoc* design to fit the situation by modifying one of the standard designs or borrowing ideas from several of them. In other words, a *standard design* does not necessarily mean that the design should be used without modification.

The designs we will examine in this chapter are classified into three groups: pre-experimental designs, true experimental designs, and quasi-experimental designs. We will use the following symbolic notation to depict these designs parsimoniously and to aid in easy comparisons across designs (this notational scheme is very similar to the schemes followed in several standard books):

---

[1] For a comprehensive treatment of experimental designs, see, for example, Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1979); Seymour Banks, *Experimentation in Marketing* (New York: McGraw-Hill, 1965).

O	Any formal observation or measurement of the dependent (or effect) variable that is made as part of the experimental study (symbols O_1, O_2, etc., will be used when two or more measurements are involved during the experiment)

Note: In many experiments one must also make formal measurements of the *independent* (or causal) variables to make sure that the experimental manipulations worked as planned. Such measurements, called ***manipulation checks*** will not be explicitly shown in the designs we will look at so as to keep the designs simple.[2]

Manipulation Check *Formal measurement of the independent (causal) variable in marketing experiments.*

X	Exposure of units participating in the study to the experimental manipulation or treatment (symbols X_1, X_2, etc., will be used when two or more experimental treatments are involved)

Note: The ordering of *O*'s and *X*'s from left to right will represent the time sequence in which they occur.

Experimental Group *Units in experiments that get exposed to the experimental treatment.*

Control Group *Units in experiments that do not get exposed to the experimental treatment.*

EG	An ***experimental group*** of units that get exposed to the experimental treatment (symbols EG_1, EG_2, etc., will be used when the experiment has more than one experimental group)
CG	A ***control group*** of units participating in the experiment but not exposed to the experimental treatment (symbols CG_1, CG_2, etc., will be used when there are multiple control groups)
(R)	Used in some designs in conjunction with the notations for experimental and control groups to indicate that units participating in the study have been *randomly assigned* to the groups (more on random assignment later)

Pre-Experimental Design *An experimental design that offers little or no control over the influence of extraneous factors and hence is not much better than a descriptive-research design when it comes to making causal inferences.*

PRE-EXPERIMENTAL DESIGNS

Studies using ***pre-experimental designs*** are, strictly speaking, not experimental studies at all. They have little or no control over the influence of extraneous factors and hence are not much better than descriptive studies when it comes to making causal inferences. While pre-experimental studies can certainly lead to hypotheses about causal relationships, one cannot have much confidence in the existence of such relationships without additional research. The label *pre*-experimental emphasizes the fact that these studies are more exploratory than conclusive as far as causal inferences are concerned.

---

[2] For further discussion of manipulation checks, see Barbara C. Perdue and John O. Summers, "Checking the Success of Manipulations in Marketing Experiments," *Journal of Marketing Research,* 23 (November 1986), pp. 317–327.

Why should we even bother to discuss pre-experimental designs when they are not really experiments? There are two good reasons for doing so. First, studies employing pre-experimental designs often form the basis for causal inferences in the real world, and one needs to be aware of their pitfalls to avoid interpreting their findings at face value. Second, comparisons with pre-experimental designs can be helpful in highlighting the merits of true experimental and quasi-experimental designs. The pre-experimental design we will look at first is the one-group, after-only design. Then we will examine the one-group, before-after design and the two-group, *ex post facto* design.

One-Group, After-Only Design

Before we examine the basic features of the one-group, after-only design, let us consider the following two situations:

Situation A. A company introduces a new brand of margarine in four test market areas and employs a very unique and revolutionary promotional campaign for it. The brand captures at least a 10% share in each market within two months after introduction. The company's management concludes that the revolutionary promotional campaign played a major role in the market share achieved by the brand.

Situation B. The president of the United States makes a television speech soliciting public support for legislation favoring prayer in public schools. A telephone survey of those who viewed the presidential speech indicates that 70% favor such legislation. The president's speech is therefore considered to have had a significant impact on the American public.

A causal inference is implied in each of these situations. However, while the inferences seem to be sound from an intuitive standpoint, they cannot be trusted entirely. To see why, let us first depict the experimental design underlying the two situations in terms of the symbols defined earlier. Both situations involve the following ***one-group after-only design***:

One-Group, After-Only Design *A pre-experimental design in which one group of units is exposed to a manipulation and then measured.*

EG *X* *O*

One group of units (EG)—the four test markets in situation A and the television audience for the president's speech in situation B—was exposed to a manipulation (*X*)—the new brand's promotional campaign in situation A and the presidential speech in situation B. Then a single measurement (*O*) was made—the brand's market share in situation A and support for school prayer legislation in situation B. A readily apparent problem with this design is the lack of control of extraneous influences. Stated differently, factors other than X may well be partly or fully responsible for the observed results (O). Another obvious limitation is the absence of any objective standard against which the results can be compared in ascertaining whether and to what extent X influenced O.

In situation A the inference that the promotional campaign was responsible for the brand's market share was quite subjective; it was based on an *assumption* that the brand's market share would have been significantly less than 10% without the promotional campaign. Similarly, in situation B the inferred effectiveness of the president's speech stemmed from a subjective notion that the percentage of respondents favoring the legislation would have been substantially less than 70% had the president's speech not been aired. The potential fallacy of such causal inferences based on data from one-group, after-only designs is emphasized by the following statement made by a marketing research director:

> I find it difficult, if not impossible, to interpret any consumer research number in a vacuum. Knowing, for example, that 61% of a sample of homemakers rated my new furniture polish "excellent" doesn't mean much to me without knowing how many gave my *old* product, or my competitor's product, a similar rating.[3]

One-Group, Before-After Design

One-Group, Before-After Design *A pre-experimental design in which one group of units is measured, exposed to a manipulation, and then measured again.*

The ***one-group, before-after design*** is an improvement over the previous design because it includes a premeasurement. The symbolic representation of this design is as follows:

$$\text{EG} \quad O_1 \quad X \quad O_2$$

The premeasurement serves as a benchmark against which the postmeasurement is compared to determine the impact of the experimental manipulation. Nevertheless, the design is still plagued by potential validity problems. Several experiments described in the preceding chapter to illustrate validity threats—the laboratory experiment to assess the impact of a new Toyota commercial on consumers' opinions about Toyota cars; the field experiment to study the impact of a special store display for Dole canned fruits on the brand's sales; the in-home test to determine the effect of a new and improved bathtub and tile cleaner on consumer preference for the product—used this design. Each of these experiments had serious internal-validity, and possibly some external-validity, problems, as was demonstrated by our discussion in the previous chapter. Potential validity threats that can become serious in studies employing the one-group, before-after design are summarized next.

---

[3] "Don't Commit Product Testing's '7 Deadly Sins,'" *Marketing News,* January 6, 1984, Section 2, p. 20.

- In laboratory experiments, which invariably involve measurement of human participants, such as the one for assessing the impact of the new Toyota commercial: maturation effect, pretesting effect, instrument variation effect, reactive bias, and pretest-manipulation interaction bias. (See Chapter 8 for definitions and illustrations of these validity threats.)
- In field experiments involving measurements of nonhuman units, such as the one for studying the effect of the special store display for Dole canned fruits: history effect, maturation effect, and mortality effect.
- In field experiments involving measurement of human respondents, such as the in-home test of the new and improved bathtub and tile cleaner: history effect, pretesting effect, instrument variation effect, mortality effect, reactive bias, and pretest-manipulation interaction bias.

Each of these experiments can also suffer from nonrepresentative-sample bias if the units chosen for the experimental study do not adequately represent the total body of units. Of course, nonrepresentative-sample bias stems from inadequate sampling and can occur in any experimental design.

In short, although the one-group, before-after design is a shade better than the one-group, after-only design, its potential validity problems make it no more than a pre-experimental design.

Two-Group, *Ex Post Facto* Design

Two-Group,* Ex Post Facto *Design *A pre-experimental design in which a manipulation is introduced among a set of units and then the set is divided into an experimental group and a control group on the basis of postmeasurement results.*

The ***two-group, ex post facto design*** has two groups of units: one exposed and the other unexposed to the experimental manipulation. There is no premeasurement, and both groups are postmeasured after the manipulation has been introduced. The symbolic representation of this design is as follows:

EG	X	O_1
CG		O_2

Recall situation B, the study to evaluate the effectiveness of the president's speech on school prayer. A major weakness of this study was the absence of a benchmark against which to compare the finding that 70% of those who viewed the television speech favored the school prayer legislation. Suppose the telephone survey conducted in this study covered a random sample of the public, including those who did not view the speech, instead of being limited to those who viewed the speech. In other words, let us assume the survey sought the opinions of those who said they viewed the speech as well as those who said they did not. Also, suppose 20% of the latter group of respondents (i.e., nonviewers) favored the school prayer legislation, in contrast to the 70% of the viewers who favored it.

Do these results imply that the president's speech was effective in swaying public opinion in favor of the school prayer legislation? It is tempting to think so since the experimental group of respondents seems to be much more favorable toward the legislation than the control group. Yielding to such a temptation can be a serious mistake, however. Although opinions were sought and compared from an exposed and unexposed group, the catch here is that the groups were *self-selected.* That is, determination of who belonged to the experimental group and who belonged to the control group was made solely on the basis of what the respondents said in the survey following the speech. The researcher had no control over the composition of the two groups and hence no way of ensuring that they had similar feelings about school prayer before the president's speech. For instance, respondents who watched the speech by and large may have shared the president's viewpoints to begin with, including those relating to school prayer. By the same token, the majority of people who did not watch the speech may have been individuals whose opinions were at odds with those of the president. Such a sharp discrepancy between the groups with respect to their prior opinions about school prayer is bound to result in serious selection effect, a major internal-validity threat.

The very term *ex post facto* (meaning, "after the fact") in the label for this design underlines its inherent weakness—namely, that the compositions of EG and CG are determined *after* the manipulation has been introduced, with no way of ensuring prior similarity between the groups. An additional problem associated with self-selection when human participants are involved is that some respondents, either intentionally or unintentionally, might state that they were exposed to the manipulation when they really had not been, or vice versa. Hence there is no guarantee that all units in EG got exposed to *X* and all units in CG did not get exposed to *X*.

Why the two-group, *ex post facto* design is a pre-experimental design should be clear by now. The results of a study using this design are no better than cross-tabulations or correlations of data obtained through descriptive-research studies. Since the researcher has no control over group compositions or over exposure or nonexposure to the manipulation, any causal inference is based solely on the postmeasurement survey. Specifically, the survey data on the dependent variable are cross-tabulated according to units that happened to be exposed to the manipulation and those that happened not to be. As we have seen several times earlier (especially in Chapter 4), cross-tabulations of descriptive-survey data can only point out associations between variables. To make causal inferences merely from evidence of associations is risky. Yet it is not uncommon to find causal inferences being made on the basis of studies using a two-group, *ex post facto* design. A case in point is the example we discussed in Chapter 8 concerning the impact of MTV on sales of a Duran Duran album in the Dallas area. Researchers and managers must be wary of such causal inferences.

TRUE EXPERIMENTAL DESIGNS

True Experimental Design *An experimental design that has built-in safeguards for controlling all, or almost all, threats to internal and external validity.*

In sharp contrast to pre-experimental designs, ***true experimental designs*** have built-in safeguards for controlling all, or almost all, threats to internal and external validity. As we will see later, the effectiveness with which various threats are neutralized by true experimental designs may still depend to some extent on the context in which they are applied. Nevertheless, these designs are generally far superior to pre-experimental designs in terms of making causal inferences with confidence.

Random Assignment *A procedure used in experimental studies to form groups of units in such a way that the group compositions can be considered to be equivalent in all respects. This term also refers to assigning experimental treatments randomly to the various groups.*

True experimental designs have two key features that enable them to exercise tight control over extraneous influences: the presence of one or more control groups and, more importantly, the ***random assignment*** of units to various experimental and control groups. Random assignment involves distributing the sample units chosen for a study to various groups on a strictly random, objective basis so that the group compositions can be considered to be equivalent before one starts the experiment. As Kidder observes:

> Random assignment controls for the influence of all the extraneous subject [sample unit] variables that you do not want to study but also do not want to hold constant because holding them constant limits the generalizability of your study. It is the defining feature of a true experiment.[4]

Matching *A procedure used in certain experimental studies to form groups of units in such a way that the group compositions are similar with respect to some specific characteristic(s).*

Since random assignment is a critical characteristic of true experiments, we should examine how it differs from an alternative approach for ensuring group equivalence, called ***matching***, which is often used. Matching involves forming groups in such a way that the composition of units is similar across groups with respect to some specific characteristic(s). To illustrate, suppose a sample of respondents is to be divided into two groups in a laboratory experiment for testing the relative effectiveness of two different commercials for a new car. Assuming that respondents' income levels, for instance, may have a bearing on how they react to the commercials, forming two matched groups involves ensuring that the two groups have similar income distributions. Any differences in reactions to the two commercials that can be attributed to income levels will then cancel out between the two groups.

A potential limitation of matching, however, is that the groups can differ on key characteristics on which they are not explicitly matched. For example, the two groups with matched income distributions may differ significantly on characteristics such as their interests and driving habits, which, in turn, may result in differential impact of the commercials, thereby lowering internal validity. Another practical limitation of matching is that data on the matching characteristics must be readily available in order to form the groups. Matching obviously will not be possible when such data are not available.

---

[4] Louise H. Kidder, *Selltiz Wrightsman and Cook's Research Methods in Social Relations* (New York: Holt, Rinehart and Winston, 1981), p. 18.

Random assignment, in contrast to matching, equalizes groups on all relevant characteristics with no special emphasis on one or a few characteristics. Furthermore, random assignment does not require any data on characteristics of the initial sample of units. Between matching and random assignment, the latter is therefore the generally recommended approach for forming equivalent groups.[5] According to Kerlinger:

> Theoretically, randomization is the only method of controlling *all* possible extraneous variables. Another way to phrase it is: if randomization has been thoroughly accomplished, then the experimental groups can be considered equal in all possible ways. . . . All other methods leave many possibilities of unequality. If we match for intelligence, we may successfully achieve statistical equality in intelligence (at least in those aspects of intelligence measured), but we may suffer from inequality in other significantly influential independent variables like aptitude, motivation, and social class.[6]

Random Sampling
An objective procedure for selecting a representative sample of units for a study.

Before we look at specific true experimental designs, three additional points concerning random assignment are noteworthy. First, *random assignment* is not synonymous with the term ***random sampling***, which we will cover in detail in Chapters 12 and 13. Kidder nicely points out the distinction between the two:

> *Random assignment* is a procedure you use after you have a sample of subjects and before you expose them to a treatment. It is a way of assigning subjects to treatments so that the groups do not differ before the treatment begins. . . . *Random sampling* is the procedure you use to *select* the subjects you will study. Random sampling serves not to equate two or more experimental groups but to make whatever subject group you study representative of a larger population. . . . Random sampling allows you to say that what you have found in [a] sample is true of [units] in the larger [body of interest]. It maximizes the external validity of research. Random assignment, on the other hand, enables you to say, "*X* caused *Y*" with some degree of certainty. It maximizes the internal validity of research.[7]

Thus random sampling is a means of controlling *nonrepresentative sample bias,* an external-validity threat. Random assignment, by contrast, minimizes *selection effect,* an internal-validity threat.

Second, *strict* random assignment may not always be desirable. Specifically, forming matched sets of units first and then randomly assigning units within each set to the various groups may be better when the number of units available for an experimental study is small. For instance, consider a field experiment to test the effectiveness of a point-of-purchase material for a brand of cookies. Let us say that eight supermarkets in a test area have agreed to participate in the experiment. Four of the eight stores are to be in an experimental group in which the point-of-purchase material will be displayed, and the rest are to form a control group. Assigning four stores each to the experimental and control groups on a *completely* random basis may be risky in this case. For instance, what if the four stores in one group turn out to be the largest of the eight? Or what if all four stores happen to be located in neighborhoods that are wealthier than those in which the other stores are located? The obvious danger here is that the skewed group compositions will lead to selection effect. Therefore a better procedure is to first form pairs of stores subjectively so that the stores within each paired set are similar on critical characteristics like size and location. This pairing results in four pairs of stores. One store from each pair is now *randomly* selected and assigned to one group (experimental or control), and the remaining store is assigned to the other group.

---

[5] For a discussion of a few other related approaches for controlling extraneous influences, see M. Venkatesan and Robert J. Holloway, *An Introduction to Marketing Experimentation: Methods, Applications and Problems* (New York: Free Press, 1971), pp. 35–37.

[6] Fred N. Kerlinger, *Foundations of Behavioral Research* (New York: Holt, Rinehart and Winston, 1973), p. 310.

[7] Kidder, *Selltiz Wrightsman and Cook's Research Methods,* p. 19.

Random assignment is not a panacea; it is merely a procedure for minimizing the odds of systematic differences between groups at the start of an experiment.[8] However, the ability of random assignment to lower those odds depends on the number of units available for assignment: The larger the initial sample size, the more successful random assignment will be in achieving equivalence across groups. Hence when the initial sample size is quite small, some suitable matching procedure followed by random assignment from within matched sets of units is better than strict or complete random assignment.

Third, in many marketing field experiments, especially those involving units such as stores or sales territories, practical constraints may make random assignments infeasible. For example, consider a firm selling industrial chemicals through two independent sales divisions in two regions of the country. The firm has a sales force of 80, divided about evenly between the two divisions. The firm's sales vice president wants to conduct a field experiment to ascertain the impact on sales of a 5% increase in the salespeople's commission rate. An ideal procedure for the field experiment is to divide the sales force of 80 *randomly* into two groups of 40 each, increase the commission rate by 5% for one group of salespeople, make no change in commission rate for the other group, and observe the sales results. Random assignment of salespeople to the experimental and control groups may not be feasible in this case, however, because of the friction and morale problems it may create within each division's sales force. A more practical and sensible thing to do is to treat salespeople in one division as the experimental group and those in the other division as the control group, especially since the two divisions are autonomous. Which of the two divisions gets to be the experimental group can still be determined randomly by, for instance, the toss of a coin. The main point of this example is that complete random assignment may not always be practical, even if a sufficiently large sample of units is available. Studies conducted under these circumstances cannot employ experimental designs that are strictly true, and hence their findings must be interpreted with caution.

---

[8] Charles Ramond, *The Art of Using Science in Marketing* (New York: Harper & Row, 1974), pp. 181–182, further discusses this point and offers several examples.

We will now discuss three specific true experimental designs: two-group, before-after design, two-group, after-only design, and four-group, six-study design.

Two-Group, Before-After Design

Two-Group, Before-After Design *A true experimental design in which an experimental group and a control group are formed by random assignment, measured before the experimental group's exposure to the manipulation, and then measured again after the exposure.*

The symbolic representation of the ***two-group, before-after design*** is as follows:

$$\begin{array}{llll} \text{EG(R)} & O_1 & X & O_2 \\ \text{CG(R)} & O_3 & & O_4 \end{array}$$

We will first examine the general features of this design and then discuss specific applications of it. Since the study units in this design are randomly assigned to EG and CG, the two groups can be considered equivalent. That is, they will be influenced by the same extraneous factors and are likely to go through similar experiences, *except* for the exposure to the experimental manipulation (X), which will occur in EG but not in CG. Therefore the difference between the pre- and postmeasurements of CG (that is, $O_4 - O_3$) should give a good indication of all extraneous influences experienced by EG. The difference between the pre- and postmeasurements of EG (that is, $O_2 - O_1$) reflects the impact of X as well as any extraneous influences. Hence the true impact of X is given by

$$(O_2 - O_1) - (O_4 - O_3).$$

This expression completely accounts for and neutralizes all validity threats *except* the following (if they happen to be present):

1. Mortality effect. Recall that the mortality effect will occur only if certain units fail to complete the entire study *and* if the units dropping out differ systematically from the ones that remain. Since EG and CG are equivalent to begin with, the experimental results may have internal validity if similar units drop out of both EG and CG; that is, if the composition of the two groups is similar at the postmeasurement stage. But if this composition is different from that of the total body of relevant units, the findings will lack external validity.

2. Reactive bias. Reactive bias occurs when study participants behave abnormally because they know they are participating in an experiment. To the extent that EG and CG will go through similar experiences, any reactive bias should influence both groups equally. Therefore the net difference between the EG and CG measurements—that is, $(O_2 - O_1) - (O_4 - O_3)$—should still provide a broad indication of whether X had any impact. However, the external validity of the findings may be questionable if reactive bias is present. Although the typical purpose of marketing experiments is to assess the *relative* effectiveness of experimental treatments—for example, Is X better than no X? Is X_1 more effective than X_2?—rather than to accurately predict the absolute impact of a treatment[9] one should be careful in generalizing the results of laboratory experiments that are prone to reactive bias.

3. Pretest-manipulation interaction bias. As we saw in Chapter 8, pretest-manipulation interaction bias would occur when the premeasurement increases or decreases the experimental units' sensitivity to the experimental manipulation. It is a potential external-validity threat and may occur in situations where human participants are premeasured concerning a topic and then are exposed to a manipulation that closely relates to the *same* topic. When the two-group, before-after design involves human participants, a premeasurement followed by a manipulation that is linked to the subject matter of the premeasurement will occur *only in EG and not in CG.* In other words, the contaminating influence of pretest-manipulation interaction will be present in the $(O_2 - O_1)$ difference but *not* in the $(O_4 - O_3)$ difference. The CG therefore does not really control for pretest-manipulation interaction bias (although it does control for *pretesting effect* wherein the premeasurement directly influences the postmeasurement). In short, the external validity of this design may be questionable when human participants are involved.

4. Nonrepresentative-sample bias. Nonrepresentative-sample bias stems from a discrepancy between the composition of the entire sample of units participating in a study and the composition of the total collection of units. Since this bias depends on the sample selection procedures rather than the nature of a particular design, it is bound to be present in *any* experiment that starts out with a nonrepresentative sample. Hence this bias, if present, is *not* due to any inherent weakness of the two-group, before-after design.

We will now look at three illustrative applications of the two-group, before-after design and identify the validity threats that will be, and those that will not be, neutralized by this design under each application.

Experiment 1. A field experiment is conducted to study the impact of a price reduction on the sales of a brand of paper towels. A sample of 50 supermarkets in a certain region is divided into EG and CG, each containing 25 stores, through random assignment. Unit sales of the paper towel brand are monitored for a four-week period in EG and CG and constitute the O_1 and O_3 measurements. Then the price of the brand is reduced 10% (that is, X is introduced) in EG but is left unchanged in CG for a four-week period. Unit sales are once again monitored during this period to yield the O_2 and O_4 measurements.

---

[9] Banks, *Experimentation in Marketing*, pp. 1–3, offers a good discussion of this point.

Experiment 2. A field experiment is conducted to measure the impact of a two-page brochure describing the harmful effects of sugar on consumption of sugar-free soft drinks. A sample of 200 households is chosen from an area. One-half of the sample is randomly assigned to EG and the other half to CG. A questionnaire about general food consumption behavior is completed by all households. A part of this questionnaire deals with current consumption levels of sugar-free soft drinks in the various households and provides the O_1 and O_3 measurements. A booklet on general nutrition and good eating habits is left behind in all households, with one key difference: The booklets given to households in EG contain the two-page brochure on the ill effects of sugar, but those given to households in CG do not. The heads of households in both groups are requested to read the booklets at their leisure. After an interval of three months the households are recontacted and requested to fill out the same questionnaire once again. The O_2 and O_4 measurements are obtained through this questionnaire.

Experiment 3. A laboratory experiment is conducted to determine the impact of a personal-computer commercial (demonstrating how easy it is to use the product) on consumer perceptions about its ease of use. One hundred consumers are recruited for this experiment and randomly divided into two groups (EG and CG). Both groups are administered a questionnaire dealing with a variety of home appliances. Through this questionnaire current perceptions about the personal computer are measured (O_1 and O_3). The EG and the CG are then shown the same hour-long television program (in different viewing rooms) containing commercials for a variety of products. The test commercial (X) is inserted in the version of the program seen by EG but not by CG. Then the relevant O_2 and O_4 measurements are obtained through the same questionnaire.

The type of validity threats present and their degree of seriousness vary somewhat across experiments 1, 2, and 3, even though all three employ the same design. Table 9.1 contains a list of the validity threats and, for each experiment, indicates which threats are controlled by the design and which are not.

As demonstrated by Table 9.1, the two-group, before-after design suffers from just one serious problem *stemming from the nature of the design itself,* namely, the problem of pretest-manipulation interaction bias in experiments 2 and 3, both involving human participants. The next design eliminates this problem.

Table 9.1 *Validity Threats in Experiments 1, 2, and 3*

Type of Threat	Experiment 1	Experiment 2	Experiment 3
History effect	May occur but is controlled	May occur but is controlled	Unlikely to occur
Maturation effect	May occur but is controlled	May occur but is controlled	May occur but is controlled
Pretesting effect	Unlikely to occur	May occur but is controlled	May occur but is controlled
Instrument variation effect	Unlikely to occur	May occur but is controlled	May occur but is controlled
Selection effect	Unlikely to occur	Unlikely to occur	Unlikely to occur
Mortality effect	May occur; whether it does depends on the nature of the units dropping out	May occur; whether it does depends on the nature of the units dropping out	Unlikely to occur
Reactive bias	Unlikely to occur	May occur; if so, the EG and CG equivalence should still preserve internal validity, although external validity may be lowered	May occur; if so, the EG and CG equivalence should still preserve internal validity, although external validity may be lowered
Pretest-manipulation interaction bias	Unlikely to occur	May occur; if so, it is *not* controlled	May occur; if so, it is *not* controlled

Note: Nonrepresentative-sample bias is not listed because it stems from inadequacies in initial sample selection and is not unique to any particular experimental design.

Two-Group, After-Only Design

Two-Group, After-Only Design *A true experimental design in which an experimental group and a control group are formed by random assignment and measured after the experimental group's exposure to the manipulation.*

As its name implies, the ***two-group, after-only design*** has two groups but does not involve any premeasurements. The two groups are formed by random assignment, following which the experimental group is exposed to the manipulation while the control group is not. The two groups are then postmeasured. Hence this design looks like the two-group, before-after design except for the absence of premeasurements, as shown next (the subscripts 2 and 4 are used for the postmeasurements merely to be consistent with the previous design and to facilitate easy comparison):

EG(R) X O_2
CG(R) O_4

The potentially troublesome bias in the two-group, before-after design—that is, the pretest-manipulation interaction bias—is absent in this design because there are no premeasurements. Furthermore, since EG and CG are considered equivalent, the two groups should go through similar experiences, except for the presence of X in EG but not in CG. The O_4 measurement therefore includes all extraneous influences, while the O_2 measurement represents the impact of X and includes the same extraneous influences. Hence the true impact of X is given by the difference $O_2 - O_4$.

Consider the three illustrative experiments described in the preceding section. They would each employ the two-group, after-only design if they did not involve any premeasurements. In such a case, what would Table 9.1 (which summarizes the validity threats under each experiment when the two-group, before-after design is used) look like? It would look the same except for the following changes: Because of the absence of formal premeasurements, the pretesting effect, the instrument variation effect, and the pretest-manipulation interaction bias would be unlikely to occur in *any* of the three experiments. Consequently, experiments 2 and 3, both involving human participants, would no longer face the threat of pretest-manipulation interaction bias. Eliminating the threat of this bias is perhaps the single most important advantage of the two-group, after-only design over the two-group, before-after design. Another obvious strength of the design is that it requires two fewer measurements than the previous design.

You are probably wondering how the true impact of X can be obtained from just two measurements (O_2 and O_4), when the previous design required four measurements and yet could not guarantee the absence of pretest-manipulation interaction bias. Stated differently, does the two-group, after-only design have any catch? Before we answer this question, let us reexamine the expression for the impact of X derived under the previous design, namely,

$$(O_2 - O_1) - (O_4 - O_3).$$

Rearranging terms, we can write this expression as

$$(O_2 - O_4) - (O_1 - O_3).$$

When the premeasurements O_1 and O_3 are the same, or when they differ by a negligible amount, this expression reduces to $O_2 - O_4$. As we saw earlier, $O_2 - O_4$ represents the impact of X in the two-group, after-only design.

The derivation of the $O_2 - O_4$ difference from the expression for the impact of X in the two-group, before-after design highlights a distinct feature of the two-group, after-only design: *It implicitly treats the premeasurements as being equal (or almost equal).* In other words, a basic assumption of the design is that EG and CG would have yielded extremely similar, if not identical, premeasurements had such measurements been made.

The answer to our earlier question of whether this design has any catch depends on the validity of this basic assumption. Of course, verifying this assumption is impossible since there are no premeasurements of the groups. Nevertheless, to the extent that random assignment is supposed to create equivalent groups, this assumption is likely to be quite sound. As Campbell and Stanley point out:

> While the pretest is a concept deeply embedded in the thinking of [researchers] . . . it is not actually essential in true experimental designs. For psychological reasons it is difficult to give up "knowing for sure" that the experimental and control groups were "equal" before the differential experimental treatment. Nonetheless, the most adequate all-purpose assurance of lack of initial biases between groups is randomization.[10]

One critical caveat concerning random assignment is worth bearing in mind, however. As we saw earlier, the ability of random assignment to achieve prior equality of groups depends on the initial sample size: The larger the pool of units available for assignment, the more successful random assignment will be in creating equivalent groups. Alternatively, the smaller the initial pool of units, the greater are the chances of obtaining dissimilar groups through random assignment. Of course, appropriate matching followed by random assignment will lower the severity of this problem. Nevertheless, the basic assumption of prior equality underlying the two-group, after-only design may still be shaky when the number of units available for participation in an experiment is relatively small. Under such circumstances, using the two-group, before-after design will at least provide an opportunity (through the O_1 and O_3 measurements) to verify whether the groups are similar to start with. But then the premeasurements will introduce the possibility of pretest-manipulation interaction bias.

Fortunately, however, marketing contexts characterized by small initial sample sizes typically involve *nonhuman* units (stores, test market areas, etc.). Since pretest-manipulation interaction bias is not a concern in experiments involving nonhuman units, the two-group, before-after design can be used in such situations. As a general rule, the two-group, before-after design is better for field experiments involving nonhuman units (e.g., experiment 1 described earlier), while the two-group, after-only design is better for field or laboratory experiments involving human respondents (e.g., experiments 2 and 3 described earlier) in which sample sizes can usually be increased sufficiently without much difficulty.[11]

---

[10] Campbell and Stanley, *Experimental and Quasi-Experimental Designs,* p. 25.

[11] Many experimental studies involving human subjects and published in the marketing literature have used the two-group, after-only design (or variations of the design, such as designs involving more than two groups to study the relative effectiveness of multiple experimental treatments). For example, one interesting after-only study involving 12 experimental-treatment groups and 1 control group was recently reported: Marjorie J. Caballero and William M. Pride, "Selected Effects of Salesperson Sex and Attractiveness in Direct Mail Advertisement," *Journal of Marketing,* 48 (Winter 1984), pp. 94–100.

Four-Group, Six-Study Design

Four-Group, Six-Study Design *A true experimental design in which two experimental groups and two control groups are formed by random assignment, one experimental group and one control group are premeasured, and all four groups are postmeasured.*

The ***four-group, six-study design*** has four randomly assigned groups—two experimental and two control—and involves six measurements.[12] It is simply a combination of the two preceding designs:

$EG_1(R) \quad O_1 \quad X \quad O_2$
$CG_1(R) \quad O_3 \qquad\;\; O_4$
$EG_2(R) \qquad\;\; X \quad O_5$
$CG_2(R) \qquad\qquad\; O_6$

As an illustration of how this design is employed, recall experiment 2 presented earlier, namely, the field experiment to assess the impact of an informational brochure on the consumption of sugar-free soft drinks. The following revised procedure will be employed in that experiment if the four-group, six-study design is used instead of the two-group, before-after design. The initial sample of 200 households will be randomly assigned to EG_1, CG_1, EG_2, and CG_2, each containing 50 households. Current consumption levels of sugar-free soft drinks in the EG_1 and CG_1 households will be measured through the food consumption questionnaire. The households in all four groups will then receive the booklet on general nutrition and good eating habits, but only the booklets received by households in EG_1 and EG_2 will contain the two-page brochure on the harmful effects of sugar. Three months later the consumption of sugar-free soft drinks by households in each of the four groups will be measured, using the same questionnaire as before.

This experiment will provide a wealth of data to measure not only the impact of the brochure (i.e., the experimental manipulation X) but also the individual magnitudes of the pretesting effect and the pretest-manipulation interaction bias. Since the four groups are formed through random assignment, premeasurements of EG_2 and CG_2, had such measurements been made, should be similar to O_1 and O_3. Hence a logical estimate of the premeasurement corresponding to EG_2 and CG_2 is the average of O_1 and O_3, that is, $(O_1 + O_3)/2$. Let D_1, D_2, D_3, and D_4 represent the differences between the premeasurement (*estimated* premeasurement for EG_2 and CG_2) and the postmeasurement for EG_1, CG_1, EG_2, and CG_2, respectively. These differences have the following interpretations:

$D_1 = O_2 - O_1$ Represents the impact of X plus all extraneous influences

$D_2 = O_4 - O_3$ Represents the impact of all extraneous influences except pretest-manipulation interaction bias

---

[12] This design was first proposed by R. L. Solomon in his article "An Extension of Control Group Design," *Psychological Bulletin,* 46 (1949), pp. 137–150; hence it is also known as the Solomon four-group design.

$D_3 = O_5 - \frac{O_1 + O_3}{2}$ Represents the impact of X plus all extraneous influences except the pretesting effect and the pretest-manipulation interaction bias

$D_4 = O_6 - \frac{O_1 + O_3}{2}$ Represents the impact of all extraneous influences except the pretesting effect and the pretest-manipulation interaction bias

The differences D_1, D_2, D_3, and D_4, given these interpretations, can be used to estimate the true impact of X, the magnitude of the pretesting effect, and the magnitude of the pretest-manipulation interaction bias. Specifically,

$D_3 - D_4 =$ true impact of X,

$D_2 - D_4 =$ magnitude of the pretesting effect,

$D_1 - D_3 =$ combined magnitude of the pretest-manipulation interaction bias and the pretesting effect,

$D_1 - D_3 - (D_2 - D_4) =$ magnitude of the pretest-manipulation interaction bias.

This design is seldom employed in marketing experiments because it is so complex and, more importantly, because marketing researchers are typically interested only in *controlling,* rather than measuring the magnitudes of, such extraneous influences as the pretesting effect and the pretest-manipulation interaction bias.[13] After all, simpler designs like the two-group, after-only design are quite capable of measuring the true impact of an experimental manipulation. Nevertheless, the four-group, six-study design is perhaps the most rigorous of true experimental designs because it offers multiple indications of a manipulation's effectiveness and increases one's ability to make causal inferences with assurance. As Campbell and Stanley state:

> The effect of X is replicated in four different fashions: $O_2 > O_1$, $O_2 > O_4$, $O_5 > O_6$, and $O_5 > O_3$. The actual instabilities of experimentation are such that if these comparisons are in agreement, the strength of the inferences is greatly increased. Another indirect contribution to the generalizability of experimental findings is also made, in that through experience with [the four-group, six study design] in any given research area one learns

---

[13] One exception is the following study using the four-group, six-study design to examine the effectiveness of corrective advertising for Listerine antiseptic: Richard W. Mizerski, Neil K. Allison, and Stephen Calvert, "A Controlled Field Study of Corrective Advertising Using Multiple Exposures and a Commercial Medium," *Journal of Marketing Research,* 17 (August 1980), pp. 341–348.

> the general likelihood of [pretest-manipulation] interactions, and thus is better able to interpret past and future [two-group, before-after designs].[14]

In short, while the four-group, six-study design may not always be practical or necessary, it is an ideal design. Experimenters may benefit by incorporating at least some of its features into any *ad hoc* designs they create to suit specific situations.

QUASI-EXPERIMENTAL DESIGNS

Quasi-Experimental Design *An experimental design that does not offer as much control as a true experimental design but usually provides more measurements and more information than a typical preexperimental design.*

Panel Design *Experimental design that uses data gathered from consumer and other types of panels.*

Quasi-experimental designs are those that do not offer as much control as true experimental designs but usually provide more measurements and more information than a typical pre-experimental design. Several quasi-experimental designs are available. However, we will focus only on one, called the time series experiment or panel design, which is the most frequently used quasi-experimental design in the marketing field.

As its name suggests, the ***panel design*** makes use of data routinely gathered from a consumer panel or other types of panels such as a panel of supermarkets. (See Chapter 4 for a discussion of panels and their capabilities and limitations.) The panel essentially acts as the experimental group, and the periodic measurements it provides become the observations (O_1, O_2, etc.). The panel design can be shown symbolically as follows:

$$\text{EG} \quad O_1 \quad O_2 \quad O_3 \quad O_4 \quad X_1 \quad O_5 \quad O_6 \quad O_7 \quad X_2 \quad O_8 \quad O_9 \quad \ldots$$

To illustrate, suppose EG is a panel of grocery stores in which sales of various brands of food products are monitored at the end of every month. Let O_1, O_2, and so on, be monthly unit sales of Oatmeal brand cereal, and let X_1, X_2, and so on, represent specific changes in Oatmeal's marketing mix that are introduced into the market area where the panel is located. Suppose X_1 is a 50-cents-off coupon good on a package of Oatmeal purchased during month 5. Exhibit 9.1 shows four hypothetical patterns of Oatmeal's sales in the panel stores during months 1 through 7 (that is, measurements O_1 through O_7).

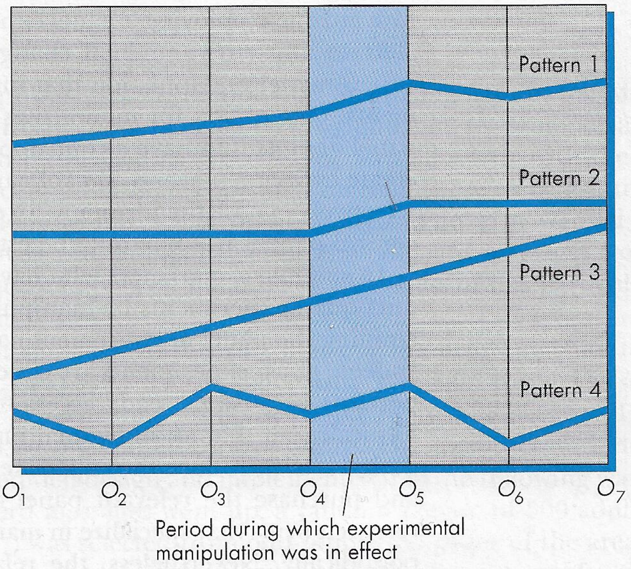
The advantage of the panel design over pre-experimental designs stems from the *pattern,* or *trend,* of measurements on the dependent variable. The trend established before introduction of the experimental manipulation—Oatmeal's sales trend during months 2, 3, and 4 in our example—serves as a sort of control in this design. A comparison of the trend after introducing the manipulation with the previously established trend can offer valuable insights that a simple one-group, before-after pre-experimental design cannot.

The shaded section of the panel design shown in Exhibit 9.1—namely, $O_4X_1O_5$—is in essence a one-group, before-after design. Assume for a moment that no measurements were made before O_4 or after O_5; that

---

[14] Campbell and Stanley, *Experimental and Quasi-Experimental Designs,* p. 25.

Exhibit 9.1 *Four hypothetical patterns of measurements in an experiment using the panel design*



is, only a one-group, before-after design was used to measure the impact of the 50-cents-off coupon on Oatmeal's sales. What would our conclusion be about the effectiveness of the coupon campaign under each of the four situations depicted in Exhibit 9.1? Looking solely at the shaded portion of the figure, we would infer that the 50-cent coupon was *equally effective* in each situation because the $O_5 - O_4$ difference is identical in all four patterns. Yet such an inference would be clearly erroneous, as discussed next.

Pattern 1 suggests that X_1 had an impact on sales but only a short-term impact; the sales trend after removal of X_1 appears to have returned to the original trend. Pattern 2 suggests that X_1 had a longer-term impact on sales, as reflected by the upward shift in the trend line even after removal of X_1. Patterns 3 and 4, in contrast, suggest that X_1 had *no* impact on sales; in each case the manipulation produced no noticeable deviation from the traditional sales pattern. These assessments of the coupon's effectiveness are much different from and *more accurate* than inferences based on the $O_5 - O_4$ difference alone.

In summary, panel designs can lead to more trustworthy causal inferences than pre-experimental designs. In addition to providing a benchmark pattern of measurements, experiments employing panels that are well designed and well run can minimize extraneous influences such as the pretesting effect, which can be serious threats in simple one-group, before-after designs.[15] However, a potentially serious drawback of panel designs is the threat of the history effect. Since panel designs are invariably used as part of field experiments, they offer virtually no control over environmental forces such as competitor's strategies. Therefore anyone employing a panel experimental design must be particularly alert to any environmental change in addition to, or instead of, the experimental manipulation that may influence the pattern of measurements. One option for minimizing the influence of the history effect in panel designs is to have a control panel (CG). Although one can seldom create experimental and control panels through random assignment and selectively expose the former to an experimental manipulation, a *matched* control panel can sometimes be established.[16]

In summary, marketers who have access to data from panels can conduct quasi-experiments by manipulating their marketing mixes and examining the panel measurement patterns to detect any significant shifts. Indeed, quasi-experiments are conducted quite frequently by marketers, perhaps because of ready access to commercially available panel data. Typically, all that an individual marketer has to do to employ a panel design is to introduce the experimental manipulation into a market area and purchase the relevant panel data from one of several marketing research firms that specialize in maintaining panels and measuring them periodically. Nevertheless, the relative ease of conducting an experimental study by using panel data should not blind one to the pitfalls of panel data, especially their susceptibility to the history effect. While quasi-experimental designs like the panel design are better than pre-experimental designs, they do not possess the rigor of true experimental designs.[17]

OTHER EXPERIMENTAL DESIGNS

As was stated at the beginning of this chapter, the standard experimental designs we have examined are not the only ones available for studying causal associations between variables. One can modify standard designs to fit a given situation by incorporating specific features and controls germane to the situation. For instance, consider the following design, which is a modified version of the two-group, before-after design.[18]

---

[15] In Chapter 4 we saw examples of well-run panels such as the National Purchase Diary Panel (NPD). Several studies have shown that sales predictions based on consumer panel data are fairly consistent with actual market results; for example, see J. H. Parfitt and B. J. K. Collins, "Use of Consumer Panels for Brand Share Prediction," *Journal of Marketing Research,* 7 (May 1970), pp. 160–167; Yoram Wind and David Lerner, "On the Measurement of Purchase Data: Surveys Versus Purchase Diaries," *Journal of Marketing Research,* 16 (February 1979), pp. 39–47.

[16] For a discussion of the use of matched panels in testing the effectiveness of promotional strategies, see Jeffrey K. McElnea, "Save Time, Money with 5-Phase Promotion Pretesting Model," *Marketing News,* May 25, 1984, Section 2, p. 2.

[17] Kidder, *Sellitz Wrightsman and Cook's Research Methods,* pp. 56–57, offers a good comparative discussion of the validity of pre-experiments, true experiments, and quasi-experiments.

$EG_1(R) \quad O_1 \quad X$

$EG_2(R) \quad \quad X \quad O_2$

This design was used by an outdoor-advertising firm in the Midwest to test the effectiveness of a billboard campaign to promote awareness about metric conversions. Specifically, ten billboards with the message "One ounce = 28.3 grams" painted on them were to be erected at various locations within a metropolitan area. The firm wanted to know whether over a three-month period the billboard messages would have any significant impact on the public's awareness of this particular metric conversion.

A major difficulty in this situation was the firm's inability to control the public's exposure to the billboards. The firm could not selectively expose an experimental group of people to the billboards and compare them with an unexposed control group. The firm therefore decided to use the modified design and implemented it in the following way: Before the billboard messages went up, a random sample of 500 adult respondents (EG_1) was selected from odd-numbered pages of the area's phone book. As part of a brief telephone survey, the percentage of respondents aware of the ounce-to-grams conversion was determined (O_1). The billboard messages were then put up (X). After three months a fresh sample of 500 respondents (EG_2) was selected from even-numbered pages of the area's phone book. This sample was interviewed, using the same telephone survey procedure as before, and the percentage of respondents aware of the metric conversion was obtained (O_2).

Since the pre- and postmeasurements came from statistically equivalent groups, the difference between them (that is, $O_2 - O_1$) indicated the impact of the billboard campaign. [In case you are wondering about the actual results, O_1 was somewhat less than 1% and O_2 was 2%. The difference between the two was not statistically significant (tests of statistical significance are discussed in Chapter 17). The billboard campaign was therefore considered ineffective.]

Note that both EG_1 and EG_2 had an opportunity for exposure to X, and EG_1 could have reacted to X with greater sensitivity than EG_2 owing to the premeasurement. However, the threat of pretest-manipulation interaction bias in EG_1 was of no concern in this study since that group's sole purpose was to provide the premeasurement that served as a benchmark.

---

[18] Campbell and Stanley, *Experimental and Quasi-Experimental Designs,* call this design the separate-sample pretest-posttest design.

Research in Use 9.1 *A. C. Nielsen Company*

Controlled Market Test

The A. C. Nielsen Company has developed ERIM TESTSIGHT, a precisely controlled testing environment that automatically measures consumer reactions to new products and marketing strategies. The TESTSIGHT service integrates in a single-source data base information gathered through electronic monitoring of grocery and drug retail activity, consumer panel purchasing, and television viewing. The TESTSIGHT markets enable a marketer to test sales performance with carefully matched cells, while maintaining complete control of variations in marketing strategies and product characteristics. And, ERIM TESTSIGHT is the only service that provides electronic insertion of test commercials into regular programming schedules, in home with or without cable. This means that a test can now be conducted using a representative sample of the entire market, not just the cable TV segment.

The compact, efficient, and controlled test market environment provided by ERIM TESTSIGHT lowers product production, media buying, consumer promotion, and trade allowance costs. A. C. Nielsen Company etimates that savings can range from more than $500,000 on an advertising test to more than $1.2 million on a new product introduction. In addition, marketing dollars are ultimately invested more efficiently because alternative strategies can be economically tested and accurately evaluated.

A NEW-PRODUCT TEST

A hypothetical new-product test case illustrates the power of the TESTSIGHT system. The purpose of the test is to evaluate the performance of a client's new brand using two introductory television ad weight plans. To implement the test, household panels in the two TESTSIGHT markets, Sioux Falls, South Dakota, and Springfield, Missouri, were divided into two cells. The cells were matched on demographic profiles, product category penetration, and volume and share levels of the major brands in the category during the months before the test brand was introduced.

A high-weight ad plan is being delivered to the entire Sioux Falls and Springfield markets, including one of the matched panel cells in each market. The second cell in each market is receiving roughly half the ad weight for the new product as the first cell. This is accomplished through the ERIM TELEMETER device, which is attached to each television in the panel households. The TELEMETER can substitute public service announcements over the air for selected advertisements, creating the effect of a low-weight ad plan in the second cell.

To control for differences between the two TESTSIGHT markets, the high-weight cells in each market have been combined to form a single

(Continued)

Research in Use 9.1 *Continued*

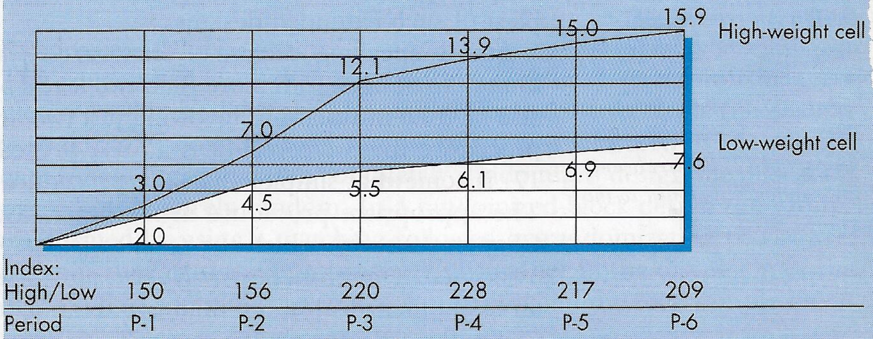
high-weight cell, and the low-weight cells have been similarly combined. Analysis of the data is being done on the basis of high-weight spending versus low-weight spending across the two markets.

THE RESULTS

After only six periods of tracking, it is already apparent that the high-weight ad plan has resulted in dramatically higher trial of the new product than the low-weight plan. Most of the difference was accumulated during the first four periods of the test, when the difference in ad weight was at its greatest.

The chart shows that cumulative household trial in the high-weight cell after six periods stands at 15.9%, more than twice the level of the low-weight cell. Given the client's goal of 16% trial for the new product by the end of its first year of availability, it would appear at this point that the low-weight plan will not allow this new brand to attain the goal set for it. Some variation of the high-weight advertising plan will be necessary.

Cumulative Household Trial Expressed as a Percent of Total Households



Source: ERIM Information Services Example.

Source: Excerpted from *The Nielsen Researcher*, November 2, 1985, a publication of the A. C. Nielsen Company, Nielsen Plaza, Northbrook, IL 60062. Used by permission.

Modified designs like the one used in the billboard campaign experiment are frequently employed by marketing researchers. A case in point is a controlled-market-test service—called ERIM TESTSIGHT (ERIM stands for Electronic Research for Insights into Marketing)—offered by the A. C. Nielsen Company. Research In Use 9.1 describes this service and illustrates it with a field experiment involving a new product. (See if you can diagram the experimental research design implied in this application.)

Some modified designs are quite sophisticated and are capable of measuring the relative effectiveness of multiple treatments within the same study (e.g., four versions of a television commercial for a product) or the influence of different levels of more than one independent variable (e.g., three different price levels in combination with two types of promotion). Such designs are sometimes called ***statistical designs*** because they require somewhat complex data analysis procedures for sorting out the separate effects of multiple independent variables and/or treatment levels. Let us now look at brief descriptions of certain commonly employed statistical designs. Chapter 19 further discusses these designs under "Analysis of Variance," which is the most frequently used procedure for analyzing data from such designs.

Statistical Design *A term used to denote an experimental design that measures the relative effectiveness of multiple levels of one or more independent variables and requires somewhat complex data analysis procedures.*

Completely Randomized Design

The various examples used in discussing the standard designs involved one experimental (causal) variable at just two levels, such as price reduction–no price reduction; commercial-no commercial; educational brochure–no brochure. We can extend the basic designs discussed earlier to include more than two levels of the experimental variable. The term ***completely randomized design*** is typically used for representing the simplest of such extended designs.

Completely Randomized Design *An experimental study design in which (1) the study units are randomly divided into as many groups as there are experimental treatments and (2) the treatments are randomly assigned to the groups.*

Strictly speaking, the use of the term *completely randomized design* is not limited to designs with more than two levels of the causal variable; it can also be used in situations involving just two levels. However, the analysis procedure for two levels (it is a *t*-test procedure, discussed in Chapter 17) is somewhat simpler than the procedure for more than two levels. The appropriate procedure in the latter case is ANOVA (discussed in Chapter 19), with which the term *completely randomized design* is usually associated. Therefore *completely randomized design* is typically used to label an experiment with more than two treatment levels.

To illustrate the design, we will consider a supermarket chain that wants to conduct a field experiment to test the relative effectiveness of three levels of point-of-purchase promotion for a certain product. The three levels are as follows:

1. No special point-of-purchase promotion (let X_1 represent this treatment level).
2. Special display of the product at its usual shelf location (X_2).
3. Special display at a high-traffic, end-of-aisle location (X_3).

The chain has selected a representative sample of 90 of its stores for the experimental study.

The use of a completely randomized design in this situation involves the following steps:

1. Randomly divide the 90 stores into three groups of 30 each.
2. Randomly assign the three treatments (X_1, X_2, X_3) to the three groups.
3. Monitor and compare product sales (O_1, O_2, O_3) in the three groups, over an appropriate period of time, to ascertain the relative effectiveness of the three promotional levels.

Using our symbolic notation, we can represent the completely randomized design in this example as follows:

$EG_1(R)$ X_1 O_1
$EG_2(R)$ X_2 O_2
$EG_3(R)$ X_3 O_3

Notice that in this example EG_1 is actually a control group since X_1 represents "*no* special point-of-purchase promotion." An explicit control group is not always necessary, however. Specifically, when the research objective is to evaluate the relative effectiveness of various experimental-treatment levels (e.g., four different commercials), rather than to compare the effectiveness of each against a no-treatment condition (e.g., no commercial), there is no need for a separate control group. Each experimental group will act as a control group for the others.

Randomized-Block Design

Randomized-Block Design *An experimental study design in which (1) the study units are divided into blocks, wherein each block is homogeneous (i.e., has units that are matched on some relevant characteristic) and contains as many units as there are treatments, and (2) the units within each block are randomly assigned to the treatments.*

Blocking Factor *A relevant characteristic of study units that serves as the basis for forming matched sets of units (called blocks) in certain experimental studies.*

The ***randomized-block design*** is basically a refined version of the completely randomized design. The main difference between the two designs is in terms of how the total sample of units is allocated to the various experimental groups. In a completely randomized design the allocation procedure is purely random. In a randomized-block design the groups are formed by using a matching-followed-by-randomization procedure, which is similar to the procedure we saw earlier in the discussion of the limitation of complete random assignment when the total sample size is small.

Using a randomized-block design in our supermarket chain example involves forming matched sets of 3 stores each on the basis of a relevant characteristic—called the ***blocking factor***—which, in addition to the experimental manipulation (that is, the three promotional levels), is likely to influence the dependent variable (that is, product sales in our example). For instance, the size of the sample stores (as measured by total sales or floor space) could have a bearing on the sales of the product, irrespective of the type of point-of-purchase promotion used. Store size is therefore a suitable blocking factor. Matched sets of stores are formed by ranking the 90 sample stores from largest to smallest (or vice versa). The first 3 stores form one set, the second 3 stores form another set, and so on. There will thus be 30 sets, or blocks, of matched stores. The stores within each block are randomly assigned to the three groups

(EG_1, EG_2, EG_3). The three groups will therefore have equal representation from each of the 30 blocks.

After the three groups are formed, the remaining steps in conducting the experiment are the same as in the case of a completely randomized design. The symbolic representation of the randomized-block design is also similar to that of the completely randomized design, except for the fact that the EGs are formed through matching followed by randomization. In forming blocks of matched units, we include as many units in each block as there are treatment levels (three in our example).

The randomized-block design increases the ability, or power, of the experiment to detect differences between the impacts of the various treatment levels by accounting for, or blocking, the influence of a key extraneous factor, such as store size in our example. Analysis of the data collected will reveal the separate effects on product sales from the treatment levels and from the blocking factor.

Latin-Square Design

Latin-Square Design
An experimental-study design in which two blocking factors are employed simultaneously. In this design the number of levels of each blocking factor must be the same as the number of treatments.

The basic principle underlying a ***Latin-square design*** is the same as that of the randomized-block design—namely, to block the effect of extraneous influences to get a sharper picture of the relative effects of the experimental treatments. But the Latin-square design permits us to block the influences of two extraneous factors rather than just one.

To illustrate this design, suppose the following two extraneous factors can influence the sales of the product in our supermarket chain example: store size and variety of products carried. One important requirement of the Latin-square design is that the number of blocks, or categories, of each extraneous factor must be the same as the number of levels of the independent variable. Therefore we cannot have 30 blocks of stores based on store size, as we had in the previous design. We can only have *three* blocks since there are three independent-variable levels (X_1,X_2,X_3). We can, for instance, divide the 90 stores into 30 large-size, 30 medium-size, and 30 small-size stores. Next, stores in each of these size categories are further divided into three subcategories on the basis of variety of products carried—for example, 10 wide-variety, 10 medium-variety, and 10 narrow-variety stores. This allocation process results in nine subgroups of stores, as shown by the following matrix:

	Product Variety		
Size	**Wide**	**Medium**	**Narrow**
Large	EG_1	EG_2	EG_3
Medium	EG_4	EG_5	EG_6
Small	EG_7	EG_8	EG_9

The three treatment levels (X_1,X_2,X_3) are assigned to the nine groups at random, subject to the condition that each treatment level is assigned once, *and only once,* to each block of the two extraneous factors. The following symbolic representation is one of several possible configurations for the resulting group assignments and measurements (the symbol R denotes random assignment of treatments to the groups, subject to the condition stated):

$$\begin{array}{lll}
\mathrm{EG}_1(\mathrm{R}) & X_1 & O_1 \\
\mathrm{EG}_2(\mathrm{R}) & X_2 & O_2 \\
\mathrm{EG}_3(\mathrm{R}) & X_3 & O_3 \\
\mathrm{EG}_4(\mathrm{R}) & X_2 & O_4 \\
\mathrm{EG}_5(\mathrm{R}) & X_3 & O_5 \\
\mathrm{EG}_6(\mathrm{R}) & X_1 & O_6 \\
\mathrm{EG}_7(\mathrm{R}) & X_3 & O_7 \\
\mathrm{EG}_8(\mathrm{R}) & X_1 & O_8 \\
\mathrm{EG}_9(\mathrm{R}) & X_2 & O_9
\end{array}$$

Analysis of sales results (the O's) will not only indicate the relative effectiveness of the three point-of-purchase promotion levels but also reveal the effects of the two blocking factors on sales of the product.

Factorial Design

Factorial Design

An experimental-study design in which two or more independent (causal) variables are manipulated simultaneously.

A ***factorial design*** involves the simultaneous manipulation of two or more independent (causal) variables. For instance, suppose the supermarket chain in our example wants to study the impact of a price reduction for the product in addition to the relative sales effectiveness of the three point-of-purchase promotion levels (X_1,X_2,X_3). For the sake of simplicity, let us assume that price has just two treatment levels: regular price (Y_1) and reduced price (Y_2).

Conceptually, the basic factorial design is like a completely randomized design with as many groups (treatment levels) as there are *combinations* of independent-variable levels. In our example, since there are three levels of point-of-purchase promotion and two levels of price, there are six combinations of independent-variable levels: X_1Y_1, X_1Y_2, X_2Y_1, X_2Y_2, X_3Y_1, and X_3Y_2. We therefore randomly divide the total sample of 90 stores into six groups of 15 stores and then randomly assign the six treatment combinations to those groups. We can depict this design symbolically as follows:

$$\begin{array}{lll}
\mathrm{EG}_1(\mathrm{R}) & \begin{matrix} X_1 \\ Y_1 \end{matrix} & O_1 \\
\mathrm{EG}_2(\mathrm{R}) & \begin{matrix} X_1 \\ Y_2 \end{matrix} & O_2
\end{array}$$

$$\begin{array}{lll} EG_3(R) & X_2 & O_3 \\ & Y_1 & \\ EG_4(R) & X_2 & O_4 \\ & Y_2 & \\ EG_5(R) & X_3 & O_5 \\ & Y_1 & \\ EG_6(R) & X_3 & O_6 \\ & Y_2 & \end{array}$$

Analysis of the data gathered through this design will provide insights into the *individual effects* on product sales of the two independent variables. For example, is the reduced price more effective than the regular price? Is the end-of-aisle promotional display more effective than the display at the usual product location? The design will also provide insights about the ***interaction effects***. An interaction effect is said to be present when the relative impact of different levels of one independent variable depends on the levels of another independent variable. For example, is the price reduction more effective when there is an end-of-aisle display than when there is no special display? Is the regular price less effective when there is no display than when there is a special display at the usual product location?

Interaction Effect *Effect in an experiment when the relative impact of different levels of one independent variable depends on the levels of another independent variable.*

SUMMARY

There are three broad classes of experimental designs: pre-experimental, true experimental, and quasi-experimental designs. Pre-experimental designs offer little or no control over extraneous factors. The least rigorous of pre-experimental designs is the one-group, after-only design. This design not only lacks control of extraneous factors but also has no explicit benchmark against which the experimental results can be compared.

An improvement over the preceding design is the one-group, before-after design, which does provide a benchmark in the form of the premeasurement. But causal inferences from studies employing the one-group, before-after design are still quite risky. This design has no control over many of the various validity threats, although which threats become serious in a situation depends on the type of experiment (laboratory or field) and the type of participants (human or nonhuman) involved.

A third pre-experimental design is the two-group, *ex post facto* design. This design provides an unexposed control group against which the experimental group can be compared. Unfortunately, the two groups are essentially self-selected and are formed *after* the experimental manipulation is introduced. The invariable result is a serious selection effect that greatly lowers one's confidence in causal inferences stemming from the use of this design. Causal inferences from studies using pre-experimental designs are tentative, at best.

True experimental designs differ from pre-experimental designs in two key ways: (1) They have one or more control groups (established before the start of an experiment), and (2) the various experimental and control groups are always formed through random assignment of the initial sample of units. Random assignment theoretically ensures equivalence among the groups on all characteristics so that any selection effect can be ruled out. It differs from an alternative procedure called matching, in which groups are formed by explicitly ensuring similarity on one or a few key characteristics.

Matching, while equalizing groups on one or a few characteristics, does not necessarily rule out dissimilarities on other characteristics. Furthermore, matching requires data on the key characteristic(s) for all the sample units. Random assignment is therefore the recommended procedure for forming experimental and control groups. One exception to this recommendation is for experiments in which the initial sample size is small. Matching followed by random assignment is better under such circumstances than strict random assignment. Also, random assignment may not always be possible (especially in field experiments) even if the initial sample is sufficiently large. Experiments using procedures other than random assignment are not *true* experiments, and their findings must be interpreted with due caution.

A commonly used true experimental design is the two-group, before-after design. This design is sound except for one major problem inherent in the design itself, namely, the threat of pretest-manipulation interaction bias when human test units are involved. The two-group, after-only design overcomes this threat since it does not involve any premeasurements. However, this design does not offer the experimenter an opportunity to verify prior equivalence of the experimental and control groups. The two-group, before-after design is generally appropriate for experiments involving nonhuman units, while the two-group, after-only design is better for experiments involving human units.

A combination of the two-group, before-after and two-group, after-only designs is called the four-group, six-study design. This design, although methodologically rigorous, is seldom used in practice because it is complex and because researchers are typically interested in only controlling, rather than measuring, extraneous influences like the pretesting effect and the pretest-manipulation interaction bias.

Quasi-experimental designs lack the strict controls inherent in true experimental designs, but they provide more information than pre-experimental designs. The only quasi-experimental design we discussed was the panel design, in which periodic observations from an established panel provide a pattern, or trend. This preestablished pattern becomes a benchmark, or control, for assessing the impact of any experimental manipulation introduced between measurements. Causal inferences from a panel design experiment are therefore more trustworthy than those from studies using pre-experimental designs. However, since panel designs are invariably employed in field experiments, a drawback of such designs is the threat of the history effect.

The various standard designs can be, and frequently are, modified to create *ad hoc* designs suitable for a given set of circumstances. Several of these modified designs are quite sophisticated and allow for simultaneous investigation of multiple treatments pertaining to one or more independent variables. Such designs, called statistical designs, include completely randomized designs, randomized-block designs, Latin-square designs, and factorial designs.

KEY TERMS

Manipulation check
Experimental group
Control group
Pre-experimental design
One-group, after-only design
One-group, before-after design
Two-group, *ex post facto* design
True experimental design
Random assignment
Matching
Random sampling
Two-group, before-after design
Two-group, after-only design
Four-group, six-study design
Quasi-experimental design
Panel design
Statistical design
Completely randomized design
Randomized-block design
Blocking factor
Latin-square design
Factorial design
Interaction effect

QUESTIONS

1. The ABC Company sells a line of packaged-food products. It recently test-marketed a new frozen dessert. This product achieved a 5% share of the frozen-dessert market in the test area within a period of six months. On the basis of this result the firm decided to introduce the frozen dessert nationally because, in the opinion of ABC's marketing manager, "New food products typically achieve only a 2% to 3% market share at the end of the year they are introduced." What specific experimental design and causal inference are implied in this scenario? What is your evaluation of the validity of the implied causal inference? Explain your answer.

2. Why is the one-group, before-after design not a true experimental design even though it has a benchmark measurement?

3. Would the internal validity of a laboratory experiment using the one-group, before-after design be higher than that of a field experiment using the same design? Explain your answer.

4. A grocery store in a certain town ran a full-page ad in the town's newspaper offering a triple-coupon deal—that is, the store would redeem manufacturers' coupons at three times their face value—to customers who shopped at the store between 7 p.m. and 11 p.m. the next day. A random telephone survey of 200 adults on the day following the triple-coupon promotional day produced these results: 100 of the 200 respondents had been to the store between 7 p.m. and 11 p.m. the previous day; of these 100 respondents 80 had seen the store's full-page ad. Of the 100 respondents who had not been to the store during the triple-coupon promotion, only 50 had seen the store's ad. On the basis of these findings the store's management considered the triple-coupon promotion a "big success." Do you agree? Explain your answer.

5. "Pre-experimental designs are helpful in generating hypotheses about causal relationships between variables." Discuss this statement.
6. What key features do true experimental designs have that pre-experimental designs lack?
7. Why is random assignment generally better than matching? Why is it not *always* better?
8. Explain the difference between *random assignment* and *random sampling*.
9. Will an experiment using the two-group, before-after design and involving human test units have high internal validity? Will it have high external validity? Explain your answer.
10. Should the two-group, after-only design be used only in laboratory experiments? Why or why not?
11. A national retail store chain has 1500 stores dispersed throughout the country. The chain's management wants to test the relative impact of two different, special in-store promotional displays (say, X_1 and X_2) on individual-store sales. It also wants to know whether not having any special in-store promotional display is necessarily worse than having X_1 or X_2. Create and describe an appropriate true experimental design capable of meeting management's information needs.
12. A four-group, six-study design was used in a laboratory experiment to measure the impact of an energy conservation commercial (X) on people's attitudes toward energy conservation. Respondents' attitudes were measured on a 100-point scale (higher numbers indicating more favorable attitudes toward conservation). Average attitude scores produced by the six measurements were as follows: $O_1 = 60$, $O_2 = 90$, $O_3 = 60$, $O_4 = 75$, $O_5 = 68$, $O_6 = 65$. Estimate the magnitudes of the true impact of X, the pretesting effect, and the pretest-manipulation interaction bias. What is your assessment of the energy conservation commercial's potential effectiveness?
13. Explain why a causal inference stemming from a panel design is more trustworthy than one stemming from a one-group, before-after design.
14. Critically evaluate the experiment described in the chapter to assess

the effectiveness of the billboard messages dealing with metric conversions. Specifically, go through the list of potential internal- and external-validity threats, and briefly discuss whether each threat would or would not have been a serious problem in this experiment.

15. A firm has designed three alternative packages for its new liquid detergent, A, B, and C. The price of the product will depend on which package is used. It will be highest if A is used and lowest if C is used; it will be in between if B is used. Propose and discuss a suitable field experiment that this firm can use to decide which package is best.

Cases for Part Three

CASE 3.1 *WTTW–CHICAGO*

CASE 3.2 *HAMILTON BANK (A)*

CASE 3.3 *SID'S SANDWICH SHOPS*

CASE 3.4 *HERSHEY CHOCOLATE COMPANY (A)*

CASE 3.5 *PROFESSIONAL SKI SUPPLY COMPANY*

CASE 3.1

WTTW–Chicago

Charles Gibson, advertising manager for a national computer software company, is wondering whether he should buy 30-second television commercial spots from WTTW, a public broadcasting (PBS) station in the Chicago area. He recently received a letter from the president of WTTW describing a unique opportunity for marketers of selected products and services to advertise to otherwise hard-to-reach audiences. Mr. Gibson's firm produces a variety of computer software packages for home and personal computers. The applications of these programs range from home entertainment (e.g., puzzles, games) to business applications (e.g., inventory and accounting systems).

TV ADVERTISING ON PBS STATIONS

Recognizing the growing difficulty faced by PBS stations in funding their programs through public donations and state/federal support, the U.S. Congress authorized a one-year experimental program (from June 1982 to June 1983) to permit limited advertising on ten public television stations across the country. WTTW–Chicago was one of the ten stations selected for this experimental program.

The materials accompanying the letter Mr. Gibson received from the president of WTTW outlined several restrictions concerning commercials that will be aired on the station. Some key restrictions are given next.

The information in this case was provided by courtesy of WTTW–Chicago. WTTW, 5400 North St. Louis Avenue, Chicago, IL 60625.

- In general, WTTW will exercise the same judgment in considering the acceptability of messages that is applied to programming decisions by the station. Messages that viewers may deem in poor taste, intrusive, or otherwise out of character with the program environment are unacceptable.
- Paid announcements intended to promote any opinion or point of view regarding any matter of public importance or interest, any political issue, or any matter relating to religion will not be accepted.
- Paid announcements intended to support or oppose any candidate for political office will not be accepted.
- Paid announcements promoting distilled liquor, cigarettes, or personal-hygiene products will not be accepted.
- WTTW reserves the right to refuse or discontinue any advertising at its sole discretion.
- No commercials will interrupt program material.
- Maximum of two commercial minutes per break.
- Maximum of four commercial minutes per broadcast hour.
- Commercials for directly competitive products or services will not be scheduled in the same program interval.

A 30-second spot on WTTW will cost $1080 during prime time (6–10 p.m. daily and 3–6 p.m. Saturdays). Network-owned stations in the Chicago area charge between $2000 and $10,000 for a 30-second prime-time spot.

THE WTTW AUDIENCE

One brochure sent by WTTW contained several pieces of information about the nature of the audience it attracted. According to a TV audience report prepared by A. C. Nielsen Company in March 1982, 69% of the Chicago area households with TVs (about 2,121,000 households) watched WTTW at least for some time during a typical week. During prime time alone, WTTW had a cumulative audience of 40%, or about 1,205,000 households, in an average week.

According to the A. C. Nielsen Metered Market Household Analysis for February 1982, covering Monday through Friday, 7–10 p.m., WTTW seemed to have a definite advantage over other Chicago area TV stations in terms of reaching households in which the head of household had an income of $40,000 or more, held a professional or managerial position, and had four or more years of college education. From the A. C. Nielsen data WTTW prepared a table (Table 1) summarizing the upscale nature of its audience.

Another table prepared by WTTW—again from the February 1982 A. C. Nielsen audience data—showed that a large proportion of those

Table 1 *WTTW Summary*

(a) Index of All-Households Average Quarter-Hour Ratings to Households Where the Household Head Has an Income of $40,000+

Station	Index	WTTW Advantage
WTTW	181	–
WBBM	98	85%
WMAQ	87	108%
WLS	91	99%
WGN	72	151%
WFLD	72	151%

(b) Index of All-Households Average Quarter-Hour Ratings to Households Where the Household Head Has a Professional or Managerial Position

Station	Index	WTTW Advantage
WTTW	141	–
WBBM	85	66%
WMAQ	86	64%
WLS	88	60%
WGN	105	34%
WFLD	103	37%

(c) Index of All-Households Average Quarter-Hour Ratings to Households Where the Household Head Had 4+ Years of College

Station	Index	WTTW Advantage
WTTW	170	–
WBBM	87	95%
WMAQ	80	113%
WLS	82	107%
WGN	103	65%
WFLD	92	85%

who generally watched very little TV were among WTTW's audience (see Table 2).

AUDIENCE REACTION TO COMMERCIALS ON PBS STATIONS

The packet of materials Mr. Gibson received also contained reprints of two articles published in August 1982. One article was titled "Preliminary

Table 2 *Index of All-Households Average Rating (100) to Households Where Viewing Is in the Lightest Quintiles, Monday–Sunday Prime Time (7 P.M.–10 P.M.)*

	Index of Viewing		
Station	**All Households**	**Lightest 20% Households**	**Lightest 40% Households**
WTTW	100	67	97
WBBM	100	35	57
WLS	100	27	54
WMAQ	100	27	53

WTTW Advantage Among Lightest 20% Viewing Households

WTTW vs. WBBM	+91%
WTTW vs. WLS	+148%
WTTW vs. WMAQ	+148%

WTTW Advantage Among Lightest 40% Viewing Households

WTTW vs. WBBM	+70%
WTTW vs. WLS	+80%
WTTW vs. WMAQ	+83%

Survey Finds Acceptance of Commercials in Public TV Programming" (*Marketing News,* August 6, 1982), and the other was titled "PBS Viewers Surprisingly Tolerant of Ads" (*Adweek,* August 9, 1982). Both articles cited a study conducted by John D. Abel, chairman of the telecommunications department at Michigan State University. A key finding of the study was that one-half to three-quarters of the more than five thousand respondents said they favored or strongly favored the use of commercials on public television as a means of raising funds for PBS stations. The survey respondents were PBS subscribers and were chosen from the ten cities taking part in the trial use of advertising on public television. The strongest support for advertising on public television came from Pittsburgh area respondents (with 74.8% favoring or strongly favoring such advertising), and the least support came from New York area respondents (49.1%). In the Chicago area 53.9% of the respondents said they favored or strongly favored airing commercials on public television.

QUESTIONS

1. Would you recommend that Mr. Gibson buy commercial time on programs aired by WTTW? Why or why not?
2. Identify the various types of secondary data that are in the packet of

materials Mr. Gibson received from WTTW. Evaluate each type of data in terms of its likely accuracy and relevance for the decision facing Mr. Gibson.

3. What other sources and types of data, if any, could or should Mr. Gibson consider before deciding whether to buy commercial time on WTTW?

CASE 3.2

Hamilton Bank (A)

Hamilton Bank, located in Lancaster, Pennsylvania, is a large bank serving a number of customers in the retail (i.e., personal/household accounts) and commercial (i.e., business accounts) sectors. In the early eighties Hamilton Bank was one of the few banks in the country that had a separate marketing research department. Hamilton Bank's marketing research department was headed by a director of market research and had several full-time employees.

In response to a request from the head of the Commercial Division of the bank, the director of market research submitted a research proposal, titled "Commercial Services Use and Image Study," on the first day of August. A description of this proposal follows.

OBJECTIVE

The purpose of this research study is to lay a foundation for a commercial marketing plan.

How is Hamilton viewed by businesses as a supplier of commercial services? In which services are we viewed as being relatively strong/relatively weak? How are we viewed in comparison with major competition? What services are being used most commonly? What services will companies be looking to banks for in the future? What are the media habits of the individuals in companies making banking decisions? The objective of this study is to provide answers to these questions.

The information in this case was provided by courtesy of Hamilton Bank, P.O. Box 3959, Lancaster, PA 17604.

METHODOLOGY

A total of 1750 mail questionnaires will be sent to businesses in our six county trade areas plus Lehigh County (250 per county). This mailing will follow a pretest of the questionnaire accomplished by conducting approximately ten personal interviews.

Before the questionnaires are sent, a telephone call will be made to each company asking the name of the person making banking decisions, informing the company of the imminent arrival of a questionnaire, and soliciting cooperation in completion of the questionnaire. Follow-up contacts may be necessary in order to solicit cooperation.

All businesses will be randomly selected from a listing presently in-house. Image results will be reported at the following levels:

1. Major SIC classifications.
2. Brand image (our customers); competitive image (noncustomers of Hamilton Bank); total image for Hamilton Bank.
3. Competition without Hamilton brand/competitive/total image.
4. Image by trade area: Hamilton, competition; brand, competitive, total.

TIME AND COST

The majority of this project will be done internally by the Market Research Department. An outside consultant will be retained to do the required statistical calculations.

The costs for this project are estimated as follows:

Internal labor	$3000
Consultants' fee	1000
Postage	700
	$4700

This project will be completed and a report issued by the end of September.

This proposal was approved by the head of the Commercial Division.

QUESTIONS

1. Evaluate the strengths and weaknesses of the proposed research design.
2. If you were the head of the Commercial Division, what questions, if any, would you raise about the proposal? What alternative approaches would you suggest?

CASE 3.3

Sid's Sandwich Shops

Sid's Sandwich Shops is a regional fast-food chain with 180 outlets spread across seven states. Last year's sales at these outlets ranged from about $100,000 to over $900,000 per outlet, depending on size and location. The average sales per outlet was around $200,000.

Sid's Sandwich Shops recently added a new breaded fish filet sandwich to its menu in a selected group of 30 outlets. This sandwich, priced at $2.69, is the most expensive item on the menu. Therefore management of the fast-food chain wanted to evaluate the potential impact of the sandwich on overall sales in a few selected outlets before deciding whether to offer the sandwich through all its outlets.

The 30 outlets in which the sandwich was introduced were selected carefully to represent the entire range of outlets in terms of sales. Specifically, all 180 outlets were arranged from smallest to largest in terms of last year's total sales volume. Then one outlet between the first and sixth outlets in this list was chosen randomly. Every sixth outlet thereafter in the list was chosen to yield a total of 30 outlets in which the fish filet sandwich was introduced. Last year's average annual sales per outlet within this group of 30 was $192,000.

MARKETING OF THE FISH FILET SANDWICH

The fish filet sandwich was introduced at the beginning of July, and monthly sales in the 30 test outlets were monitored through the end of December. So that customers' attention was directed to the fish filet sandwich, special point-of-purchase promotional displays were set up in the outlets selling the sandwich during the six-month period. Messages such as "Try our new, tasty fish filet sandwich" were prominently displayed throughout the outlets.

Sid's Sandwich Shops did not promote the fish filet sandwich through any means other than the point-of-purchase messages displayed at the 30 outlets. The primary reason for not promoting the new product through media such as newspapers, radio, and television was that the sandwich was not available in all outlets of the chain. Because of the random selection of the 30 outlets, in the same general area one outlet might be offering the sandwich and another might not. Promoting the sandwich through the mass media was therefore considered to be inappropriate.

SALES IMPACT OF THE SANDWICH

The average monthly sales during the six-month period (July through December) for the test outlets was approximately $16,000 per outlet. The average monthly sales for these same outlets during the six months preceding the introduction of the fish filet sandwich (i.e., January through July) had been about $16,300. Sid's Sandwich Shops' management was quite disappointed with these results. For comparison purposes management also examined the first six-month and second six-month average monthly sales for the 30 test outlets during the three preceding years. These sales figures are as follows:

	Jan.–June	**July–Dec.**
One year ago	$16,200	$15,800
Two years ago	$15,750	$15,300
Three years ago	$15,950	$15,550

After comparing this year's sales in the test outlets with the sales figures tabulated above, management concluded that the fish filet sandwich "was not a viable new product capable of significantly adding to our chain's sales." They therefore decided to withdraw the product.

QUESTIONS

1. Do you agree with the decision made by Sid's Sandwich Shops' management? Why or why not?
2. Would you characterize the research implied in this situation as an

experiment? If so, what type of experimental design was used? Explain your answer.

3. If you were to start over again to assess the sales impact of the fish filet sandwich, what would you do differently? Why?

CASE 3.4

Hershey Chocolate Company (A)

Hershey Chocolate Company is a division of Hershey Foods Corporation, headquartered in Hershey, Pennsylvania. Hershey Foods Corporation had over 13,500 full-time employees and net sales of over $1.5 billion in 1982. In addition to producing a wide variety of chocolate and confectionary products, the firm is a major producer of pasta products and operates a chain of restaurants as well as an office coffee service plan.

ALTERNATIVE RECIPE STUDY

In early 1982 the Hershey Chocolate Company was in the process of designing a magazine advertisement for one of its products, namely, peanut butter baking chips. A key question was which, if any, of three alternative recipes utilizing the peanut butter baking chips should be used in the magazine advertisement. The three recipes were for a regular cookie, a bar cookie, and a peanut-flavored brownie.

A study involving 200 shopping mall intercept interviews with female heads of household was conducted from March 31 through April 16, 1982, to measure consumers' initial interest in and reactions to trials among the three recipes. One-half of the mall intercept respondents were current users of Hershey's peanut butter chips, and the other half were scratch bakers whose families liked peanut butter–flavored desserts. The respondents were between the ages of 18 and 49 and had at least one child under 17 years of age living at home.

The information in this case was provided by courtesy of Hershey Foods Corporation, 19 East Chocolate Avenue, Hershey, PA 17033.

Table 1 *Degree of Appeal of Recipe Boards (in Percentages)*

	Board Number		
Response	**231 (Regular Cookie)**	**479 (Bar Cookie)**	**658 (Brownie)**
Very appealing	79	52	66
Somewhat appealing	18	33	22
Neither appealing nor unappealing	2	8	5
Somewhat unappealing	0	4	4
Very unappealing	0	2	2
Total number of respondents	200	200	200

Note: Percentages may not add to exactly 100% owing to rounding.

Each of the three recipes was mounted on a board along with a picture of the dessert the recipe was intended for. Each board was assigned an identification number: 231 for the regular cookie recipe board, 479 for the bar cookie recipe board, and 658 for the brownie recipe board.

During the interview the three boards were handed one at a time (e.g., 231 first, 479 second, and 658 last) to the respondent. The ordering was rotated from respondent to respondent to avoid a positioning bias for any of the three recipes. After looking at each board, the respondent indicated her reactions to the picture and the recipe on it. The respondents were asked how much the dessert in the picture appealed to them in terms of being appetizing and how likely they were to try the recipe if they happened to see it in a magazine.

After reacting to each board separately, the respondents were shown all three boards together and asked which recipe they would be most likely to use. The respondents were then given a copy of the recipe they were most likely to use, along with a sample of peanut butter chips and a $5.00 incentive, and were requested to try the recipe at home within the next few days.

A telephone callback interview was conducted with 191 of the original 200 respondents who had tried the recipe within the next several days. The primary purpose of this telephone survey was to measure the respondents' degree of satisfaction after using the recipe they had selected.

RESULTS OF THE RECIPE STUDY

Respondents' reactions to the pictures in the three recipe boards are summarized in Table 1. Table 2 summarizes the likelihood of respondents' making each type of recipe if they were to see it in a magazine.

Table 2 *Likelihood of Trying Recipe (in Percentages)*

	Board Number		
Response	**231 (Regular Cookie)**	**479 (Bar Cookie)**	**658 (Brownie)**
Very likely	61	39	56
Somewhat likely	28	36	22
Neither likely nor unlikely	4	4	6
Somewhat unlikely	2	7	6
Very unlikely	3	13	8
Total number of respondents	200	200	200

Note: Percentages may not add to exactly 100% owing to rounding.

Table 3 *Customer Experiences and Reactions*

Type of Recipe	**Regular Cookie**	**Bar Cookie**	**Brownie**
Number of Respondents Trying Recipe	89	36	66
Ease of Preparing*			
Very easy to make	88	86	85
Somewhat easy to make	11	11	14
Somewhat difficult to make	1	3	2
Taste of Finished Product*			
Better than expected	46	28	35
About the same as expected	43	56	50
Not as good as expected	11	17	15
Likelihood of Making Again*			
Definitely will make again	56	47	39
Probably will make again	28	33	30
May or may not make again	13	11	6
Probably will not make again	2	3	2
Definitely will not make again	0	6	3

*Numbers represent percentages of respondents giving each response; they may not add to exactly 100% owing to rounding.

When asking to select the recipe they would most likely try, 94 (47%) of the 200 respondents chose the regular cookie recipe, 36 (18%) chose the bar cookie recipe, and 70 (35%) chose the brownie recipe. Of the 94 respondents selecting the regular cookie recipe, 89 (95%) had tried it within a few days. All 36 respondents selecting the bar cookie recipe had tried it; 66 (94%) of the 70 respondents selecting the brownie recipe had tried it. Selected results concerning respondents' experiences with the recipes and their reactions to the dessert are summarized in Table 3.

QUESTIONS

1. Critically evaluate the research design and data collection procedure used in the alternative-recipe study.
2. Do you consider the study an experimental study? If so, how would you describe it in terms of the various types of experimental designs we have discussed? If not, why not?
3. From the results in Tables 1, 2, and 3, what inference(s) can you make? Why?

CASE 3.5

Professional Ski Supply Company

Professional Ski Supply Company is the manufacturer of Skiwell brand of ski clothing and equipment. The Skiwell product line includes over twenty different products that are sold through major sporting goods stores across the country. The company grossed about $15 million in sales during the past year. The company's sales growth has paralleled that of the industry as a whole. The size of the ski apparel and equipment market in the United States is estimated to be over $1 billion. However, the market is quite fragmented, with few companies enjoying annual sales of over $50 million.

Harold Sanchez, president of Professional Ski Supply Company, is wondering whether his company should employ a promotional strategy that is being increasingly used by its competitors—namely, signing up popular skiers (e.g., potential and past Olympic gold-medal winners) to wear a certain brand of apparel or use a certain brand of equipment when they ski. Mr. Sanchez has heard that some companies apparently achieved tremendous sales success by gaining the endorsement of well-known skiers. However, he did not have any concrete evidence about the cost versus the benefits of using professional skiers to promote ski gear in general and the Skiwell brand in particular.

Mr. Sanchez recently read an article in *Time* magazine dealing with the practice of using popular skiers to serve as "skiing billboards." This article gave Mr. Sanchez a mixed impression about the potential effectiveness of inducing well-known skiers to endorse certain brands of ski gear. According to one industry official, such promotions are particularly effective among a growing number of skiers who are taking up slaloming on weekends. In this official's opinion: "These are fanatic skiers. They are buying nearly anything that makes them look fast."

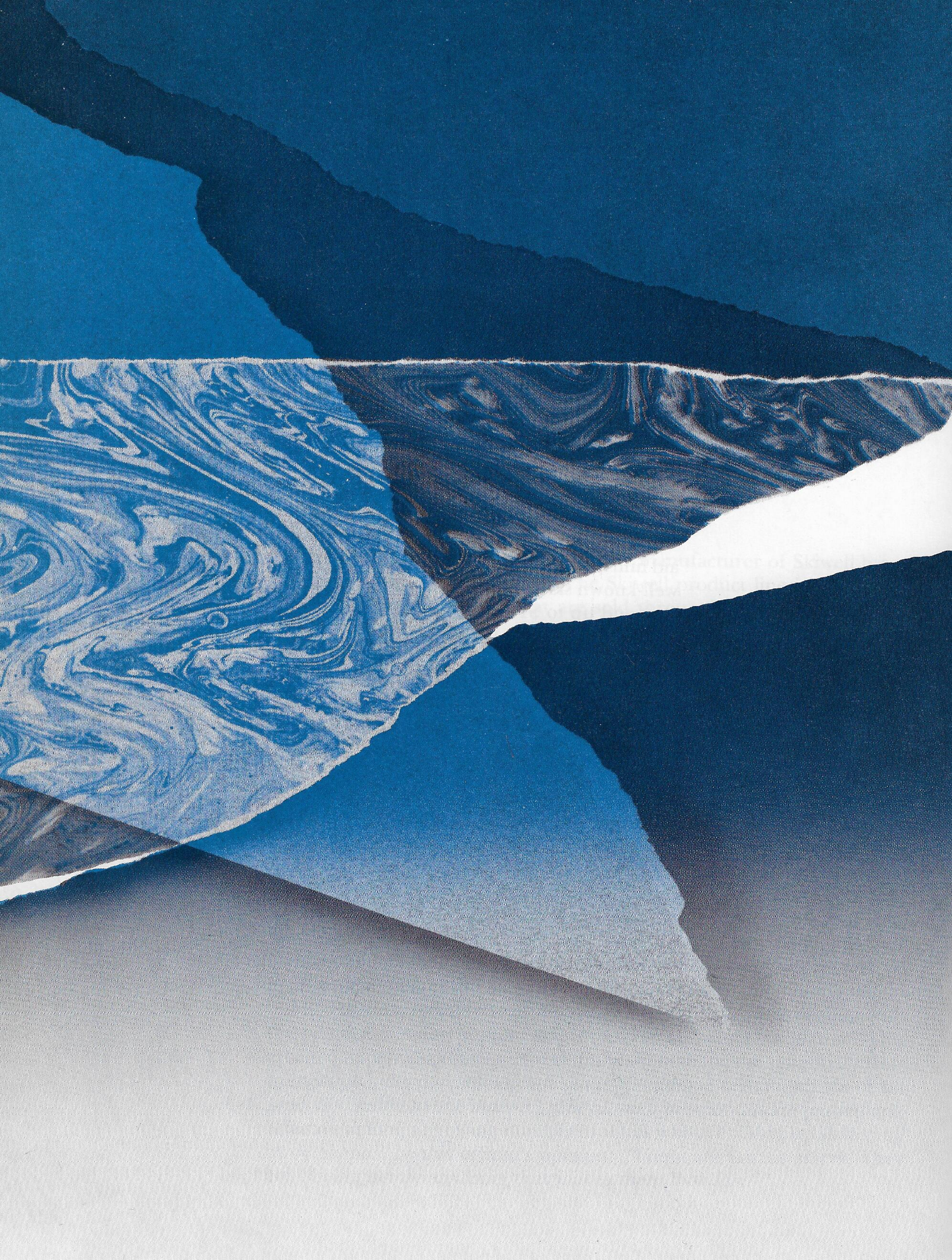
Others in the ski industry were somewhat less emphatic about the impact of promotions using well-known skiers. For instance, the president of K2, a ski maker whose brand name was being promoted by Phil and Steve Mahre (twin brothers who were top skiers in the United States), had this to say: "We won't get a sales jump just because Phil or Steve wins a gold medal. But we think our involvement with top racers gives us a lot of credibility." An executive of another leading ski apparel company was cited as saying "Most consumers are sophisticated enough to know that Phil Mahre could win in any sweater or pair of pants."

In addition to being unsure of whether Skiwell products should be promoted by using well-known skiers, Mr. Sanchez is concerned about the effort and the expense involved in obtaining the services of top racers. The competition for the services of these athletes is intense among firms in the ski industry. Top names are reportedly being paid several hundred thousand dollars to endorse a particular brand of ski wear or equipment. For instance, one ski equipment company was said to have paid a leading skier $700,000 to switch from a competitor's brand of skies to its own brand.

Mr. Sanchez wants to obtain relevant and trustworthy data that will aid him in assessing the feasibility and the potential effectiveness of using well-known skiers to endorse one or more Skiwell products. He is willing to spend up to $10,000 for generating the needed data.

QUESTIONS

1. If you are hired to conduct research aimed at Mr. Sanchez's problem, what specific types of data will you look for? Provide a rationale for each type of data that you feel should be collected.
2. What research approaches and specific techniques will you use to collect the types of data identified in Question 1? Why?



P A R T   F O U R

DESIGNING Forms & Scales for COLLECTING DATA

CHAPTER 10 *Designing Data Collection Instruments*

CHAPTER 11 *Types of Scales and Attitude Measurement*

CASES FOR PART FOUR

CHAPTER 10

Designing Data Collection Instruments

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Identify the ways in which lack of adequate attention to questionnaire design may result in inaccurate communication between the researcher and respondent.*
- *State the major components of the questionnaire design process and illustrate their interrelationships.*
- *Distinguish between structured and nonstructured questions and discuss the various options or decisions pertaining to constructing structured questions.*
- *Illustrate the various types of "problem" questions that may stem from improper wording.*
- *State and discuss several broad guidelines for properly arranging questions.*
- *Discuss why the appearance and layout of a questionnaire are important.*
- *Summarize the guidelines that should be followed to maximize the benefits of pretesting a questionnaire.*
- *State the attractive questionnaire features offered by computerized interviewing.*
- *Design effective cover letters for mail surveys and appropriate openers for personal and telephone interviews.*
- *Design an appropriate questionnaire or observation form, given the data requirements of a decision maker.*

As we saw in Chapter 6, questioning and observation methods are the two basic approaches available for gathering primary data. Of the two, the former is used much more frequently in research projects involving primary-data collection. Furthermore, although data collection instruments are needed in both methods, designing an instrument for a questionnaire study is far more difficult than designing one for an observation study. Researchers as well as users of marketing research must develop an appreciation for the difficulties in constructing a good questionnaire and the potential dangers of using a poorly designed one. The primary focus of this chapter will therefore be on designing questionnaires. Constructing observation forms will be discussed briefly at the end of the chapter.

QUESTIONNAIRE DESIGN

Questionnaire
An instrument used for eliciting and recording responses in many, but not all, research projects employing the questioning approach.

A ***questionnaire*** is simply a set of questions designed to generate the data necessary for accomplishing a research project's objectives. As was demonstrated in earlier chapters, primary data may be needed in exploratory- as well as conclusive-research projects. A key difference in data collection in these two types of projects centers around the formality and flexibility of the methods used. Primary-data collection in exploratory-research projects is accomplished in an informal, flexible fashion. Indeed, seldom is a standard questionnaire needed in such projects; rather, all that is normally required is a checklist of items to be investigated. Therefore questionnaire design considerations are not as germane to exploratory-research projects as they are to conclusive-research projects. The bulk of the material to be covered here relates to constructing standard questionnaires (i.e., those to be used in conclusive-research projects), although some of the guidelines discussed may be relevant for exploratory questioning as well.

COMPLEXITY OF QUESTIONNAIRE DESIGN

Designing questionnaires may appear to be simple, especially to those who have not designed one before. After all, you may think, once you have a clear notion of the information desired, it should be easy to formulate appropriate questions and arrange them in the form of an instrument. But experienced researchers will quickly point out that nothing is further from the truth. Indeed, even the age-old adage "Practice makes perfect" is debatable when it comes to designing questionnaires. Perhaps "Practice makes *almost* perfect" is the best a researcher can hope for in questionnaire construction.

There are no rules that if faithfully followed can guarantee a flawless questionnaire. As numerous illustrations in this chapter will demonstrate, even questionnaires constructed by skilled researchers may have drawbacks. For example, consider the following question, which appeared in a mail questionnaire the author received from an apparently experienced researcher:

> Do you consider the many marketing research texts adequate for most of your business majors at the *undergraduate* level?
>
> ________ Yes ________ No
>
> If no, briefly, why not? ________________________________

At first glance there appears to be nothing wrong with this question. However, a closer examination reveals potential problems. It is subject to varying interpretations by different respondents and may be difficult to answer meaningfully. For instance, what exactly do the words *many, adequate,* and *most* refer to in this question? Will one respondent's interpretation of *many* (or *adequate* or *most*) be the same as that of another? If Jane Smith feels there are three "adequate" texts, should she answer yes or no? What if she feels there are six "adequate" texts? Several issues like these can be raised concerning the meaningfulness of this question.

One positive feature of the question is the inclusion of "If no, briefly, why not?" This part at least offers the respondents an opportunity for an open-ended answer if they are unsure of their response. However, open-ended answers, in addition to creating potential coding problems, may just point out the difficulty the respondents had in answering the question rather than provide data directly pertaining to the purpose of the research.

The intent of this illustration is not to criticize the researcher who wrote the question but to emphasize the complexity of designing questionnaires. Generally, an outsider can easily find some fault with even the most carefully thought-out question. Yet the same outsider may be at a loss to construct a *flawless* question to replace the one being criticized. Even seasoned researchers may discover questionnaire flaws only *after* data collection occurs. This point is highlighted by the following statement made by Al Blankenship, a well-respected researcher, concerning one key question in a study that he himself conducted: "Hindsight suggests that it might have been wiser to have phrased the question in such a manner as to discriminate between characteristics necessary for success in almost any business endeavor and those crucial to marketing research."[1]

In summary, even carefully designed questionnaires are not immune to questionnaire errors. So a research project in which questionnaire design is taken lightly will most likely be worthless. As the next section shows, the questionnaire is a critical determinant of data accuracy—that is, the extent to which the collected data are error-free and trustworthy.

QUESTIONNAIRE'S IMPACT ON DATA ACCURACY

The questionnaire is the main channel through which data are obtained from respondents and transferred to researchers in conclusive-research projects employing personal interviews, mail surveys, or telephone surveys. This channel has a dual communication role: (1) It must communicate to the respondent what the researcher is asking for, and (2) it must communicate to the researcher what the respondent has to say. The accuracy of data gathered through questionnaires will be greatly influenced by the amount of distortion, or "noise," that occurs in the two types of communication. Unless a set of questions faithfully reflects a researcher's data requirements, and unless those questions are interpreted and answered correctly by respondents, the accuracy of the collected data will suffer. A sloppy questionnaire can lead to a great deal of distortion in the communication from researcher to respondents and vice versa.

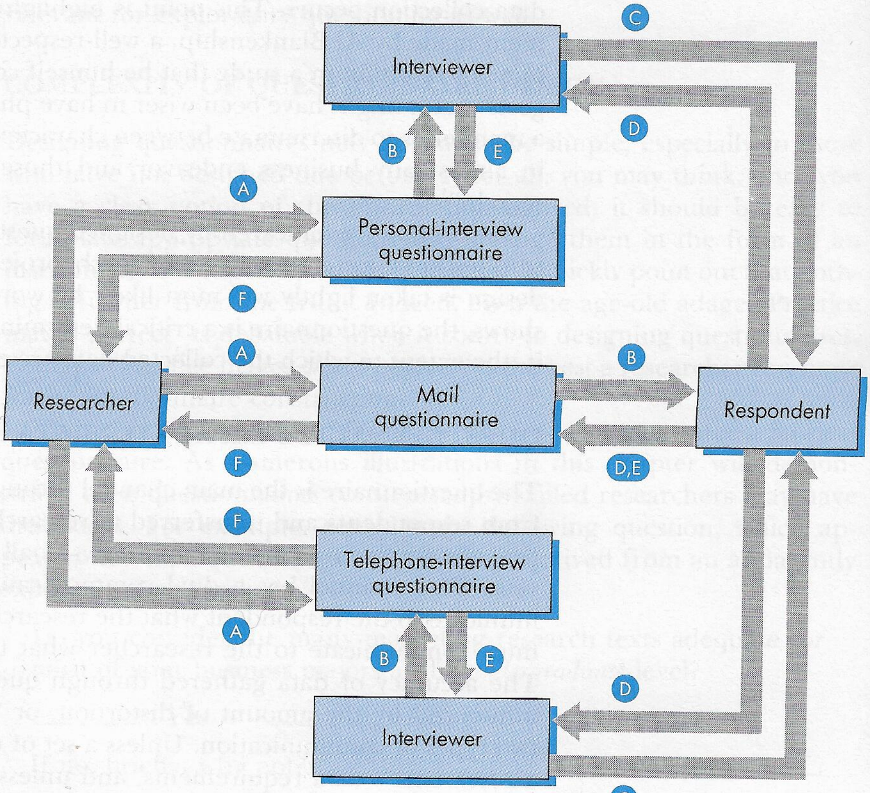
An additional source of potential distortion in questionnaire studies using face-to-face or telephone interviews is the entry of an intermediary—namely, the interviewer—into the communication channel between researcher and respondent. A poorly designed questionnaire, such as one that confuses an interviewer or is subject to varying interpretations by different interviewers, is an open invitation for the interviewer to bias the data being collected.

Exhibit 10.1 shows the vital position occupied by the questionnaire in the link between researcher and respondent. The communication flows represented by the arrows in this figure are of major concern when one is constructing questionnaires. Lack of adequate attention to questionnaire design will lead to distortions in these communication flows, thereby generating data of doubtful accuracy.

---

[1] A. B. Blankenship, "What Marketing Research Managers Want in Trainees," *Journal of Advertising Research,* 1 (February 1975), p. 13.

Exhibit 10.1 *Questionnaire—Link Between Researcher and Respondent*


C
Interviewer
D
B
E
A
Personal-interview questionnaire
F
A
B
Researcher
Mail questionnaire
Respondent
F
D,E
F
Telephone-interview questionnaire
A
B
E
D
Interviewer
C

KEY

- Ⓐ Translating data requirements into a set of questions along with instructions
- Ⓑ Interpreting the questions and instructions
- Ⓒ Administering the questionnaire
- Ⓓ Providing responses
- Ⓔ Recording the responses
- Ⓕ Interpreting the responses

Although there are no rules for developing a flawless questionnaire, the collective experience of numerous researchers offers a broad set of guidelines for minimizing the likelihood and the severity of data validity problems. Several excellent books have been written on the subject of questionnaire design.[2] Therefore we need not and cannot cover all the issues of questionnaire design in a mere chapter. Rather, the focus here will be on discussing, with several illustrations, the key tasks a researcher must perform in order to design a good questionnaire.

---

[2] One of the classic books on questionnaire design is Stanley L. Payne, *The Art of Asking Questions* (Princeton, N.J.: Princeton University Press, 1951); a more recent book on the subject is Seymour Sudman and Norman M. Bradburn, *Asking Questions* (San Francisco: Jossey-Bass, 1982).

QUESTIONNAIRE DESIGN ELEMENTS

The process of drawing up a questionnaire is a sequence of interrelated tasks. This section will provide an overview of the process; subsequent sections will describe each task in detail. Exhibit 10.2 shows the key questionnaire design tasks and the relationships among them.

The logical starting point for constructing a questionnaire is to translate the data requirements of a project into a set of rough questions. Of course, as pointed out in earlier chapters, an important prerequisite for being able to identify the appropriate types of data to be collected is a clear and correct definition of the research problem, objectives, or hypotheses. Next, certain critical checks of the rough draft have to be made: Does each question have the most appropriate *form* (e.g., *structured* versus *nonstructured*)? Is each question *relevant* and *properly worded* to obtain meaningful, valid responses? Is the *sequencing* of the questions likely to introduce any bias? Are the *layout* and *appearance* of the questionnaire conducive to accurate and easy data collection?

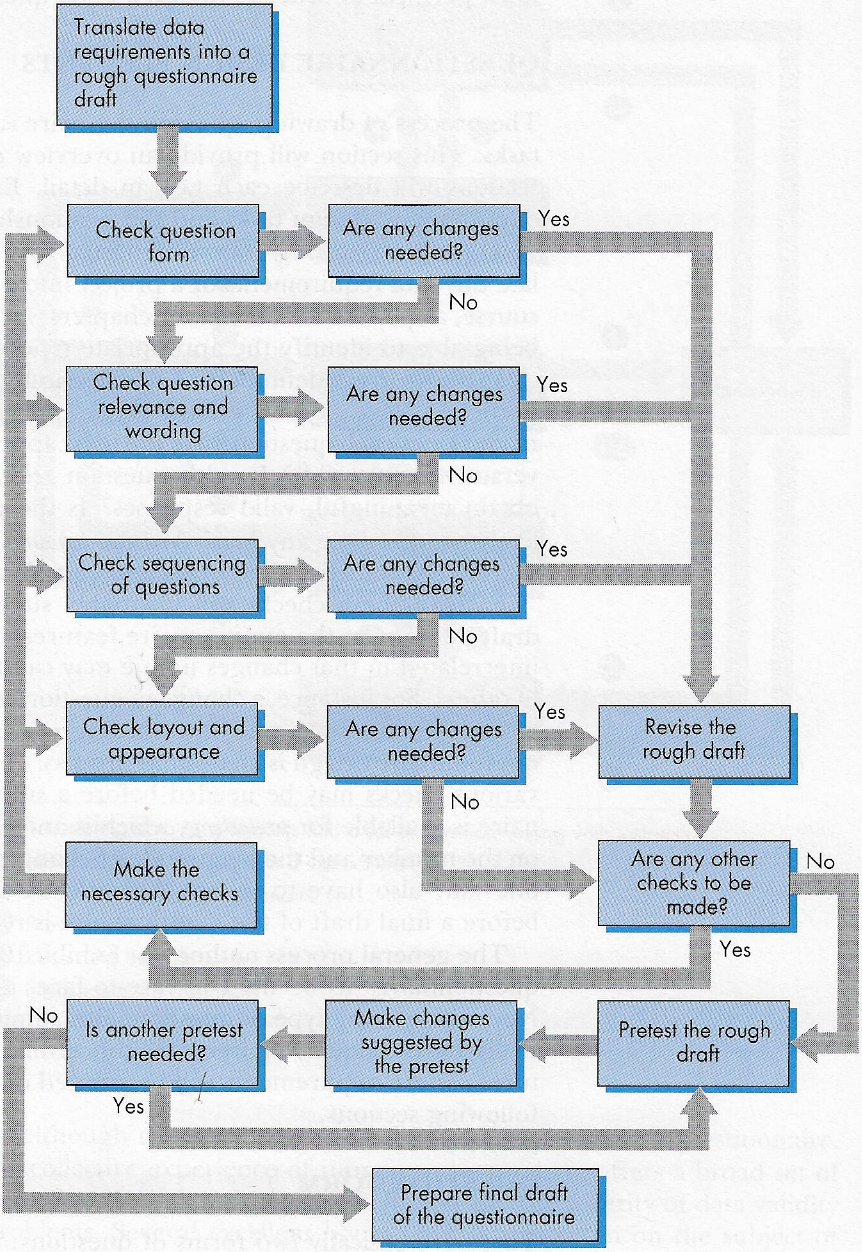
Each of these checks will invariably suggest changes in the rough draft. Moreover, the questionnaire features involved in these checks are interrelated in that changes in one may call for corresponding changes in others. For instance, a change in question sequencing may also require changes in the form and the wording of questions. The point is that questionnaire design is an *iterative* process: Numerous loops through the various checks may be needed before a suitable draft of the questionnaire is available for *pretesting*, which is another critical task. Depending on the number and the magnitude of changes resulting from pretesting, one may also have to repeat this task—in some cases several times—before a final draft of the questionnaire is ready.

The general process outlined in Exhibit 10.2 is relevant for designing questionnaires to be used in face-to-face, telephone, or mail surveys. Nevertheless, the type of questionnaire administration method will normally have unique requirements concerning certain questionnaire features. These requirements will be pointed out where appropriate in the following sections.

QUESTION FORM

There are basically two forms of questions: *nonstructured* (open-ended) *questions* and *structured* (fixed-response) *questions*. The relative advantages and disadvantages of these two question forms were covered in Chapter 6 (see Table 6.1). The emphasis here will be on looking at *variations* of the two basic forms and discussing when and how they can be most effectively used in a questionnaire.

Exhibit 10.2 *Components of the Questionnaire Design Process*


Translate data requirements into a rough questionnaire draft
Check question form
Are any changes needed?
Yes
No
Check question relevance and wording
Are any changes needed?
Yes
No
Check sequencing of questions
Are any changes needed?
Yes
No
Check layout and appearance
Are any changes needed?
Yes
No
Revise the rough draft
Make the necessary checks
Are any other checks to be made?
No
Yes
No
Is another pretest needed?
Make changes suggested by the pretest
Pretest the rough draft
Yes
Prepare final draft of the questionnaire

Nonstructured Questions

While nonstructured questions permit free responses, not all of them may require lengthy or wordy responses. Consider the following illustrative questions:

How old are you? ________

What do you like *most* about owning your own home? ________

Will you please describe your thoughts about a person who shoplifts items from a grocery store to keep from going hungry? ______

____

All three questions are nonstructured. However, the effort, time, and space required to answer them obviously vary considerably. Questions like the last one (i.e., the one dealing with shoplifting) may be inappropriate for questionnaires to be used in conclusive-research projects; the large samples typically required in such projects will make data collection, coding, and analysis particularly difficult. But open-ended questions such as the first two, especially the age question, can be effectively used even in large-scale studies since responses to them will be relatively short. Hence not *all* open-ended questions must be avoided in a standard questionnaire intended for a large sample of respondents.

Depending on the types of information desired from a conclusive-research project, one may at times need to leave certain questions open-ended, provided that the expected responses are likely to be brief. For example, compare the data to be obtained by asking the open-ended age question we just saw with data to be obtained through a structured question like this one:

In which of the following categories does your age fall?

________ Less than 18
________ 18 to 30
________ 31 to 45
________ 46 to 60
________ Over 60

The open-ended question will generate refined data (i.e., the exact ages of respondents), while the structured question will only give a rough indication of respondents' ages. Depending on the types of analyses needed to provide the information desired in a project, refined data and hence appropriate open-ended questions may be necessary. More will be said in the next chapter and in Chapters 16–19 about different types of data and their impact on the choice of analysis procedures.

Some researchers also recommend starting a face-to-face or telephone interview with a *general* open-ended question in order to put the respondent at ease and kindle interest in the study. For example, an interview dealing with home computers might begin with a question like

"What are your feelings about how the numerous electronic products available today are affecting people's life-styles?" After whatever answer is given, the respondent can be introduced to the main questionnaire with a statement such as "That is very interesting. Our interview deals with home computers. . . ."

Beginning an interview with an open-ended question is worthwhile if it will be easy to answer and aid in smooth completion of the rest of the interview. Sometimes, a general open-ended question may also be helpful during an interview to ensure smooth transition from one major topic to another (e.g., from purchase to use of a product). For such open-ended questions to accomplish their intended purpose, they must be chosen carefully, ensuring that they are unlikely to be difficult for or embarrassing to the respondent. Assuming a suitable nonstructured question is available for starting an interview, one can use it without worrying about constraints that would normally preclude its use in a large-scale survey. That is, neither coding or analyzing of the responses to the question is necessary.

Structured Questions

Structured questions are of two basic forms: dichotomous and multiple-category. A ***dichotomous question*** offers just two answer choices, typically in a yes/no format:

Dichotomous Question *A structured question that offers just two answer choices.*

Do you smoke cigarettes?

______ Yes ______ No

Have you ever watched the "Tonight Show"?

______ Yes ______ No

Questions with more than two answer choices are ***multiple-category questions***. A number of variations of the multiple-category question are possible.[3]

Multiple-Category Question *A structured question that offers more than two fixed-response choices.*

Consider the following illustrative questions:

Approximately how many long-distance telephone calls do you make per week?

______ 0 to 1 call

______ 2 to 3 calls

______ 4 to 5 calls

______ 6 to 7 calls

______ More than 7 calls

---

[3] An especially good discussion of multiple-category questions is provided by Don A. Dillman, *Mail and Telephone Surveys: The Total Design Method* (New York: Wiley, 1978), pp. 86–95.

In your opinion, which division of General Motors makes the most attractive automobiles?

________ Buick Division
________ Cadillac Division
________ Chevrolet Division
________ Pontiac Division
________ Oldsmobile Division

What do you like about First City Bank's automatic-teller machines?

________ 24-hour service
________ Privacy during transaction
________ Convenient location
________ No long lines
________ Other ______________________________
(Please specify)

Although these questions are all multiple-category questions, they differ in distinct ways that have important implications for response quality. We will discuss the differences and their implications under the following headings: response category sequence, response category content, and number of response categories.

Response Category Sequence. The response categories in the question about long-distance telephone calls follow a *natural sequence*. In fact, no alternative sequence, other than presenting the categories in reverse order, will be meaningful. However, for the other two questions the specific response categories under each can be presented in any order. This distinction is important because the sequence of presentation, by itself, can influence a respondent's choice, especially when the respondent is not quite sure how to answer the question. Researchers have found, for instance, that categories in the middle have greater drawing power than the ones at the extremes if the categories involve numbers (e.g., the question on long-distance telephone calls). When the categories are in the form of words, phrases, or statements, the categories at the extremes have greater drawing power.[4]

Response Category Sequence Rotation
The process of varying from respondent to respondent the order in which response choices to a multiple-category question are presented.

To eliminate entirely the biases stemming from response category sequence may not be possible. However, some precaution can be taken to minimize such biases. An approach frequently followed is ***response category sequence rotation***. This approach involves rotating the sequence of categories from one questionnaire to the next. Through systematic rotation of the order of presentation of response choices, each of them will occupy various positions within the sequence. The assumption underlying this procedure is that any possible response bias arising from using a particular sequence will be neutralized across all respondents.[5]

---

[4] Payne, *The Art of Asking Questions*, p. 134.

The ease and effectiveness with which sequence rotation can be accomplished depends on the nature of the response choices. As we just saw, the only sequence rotation possible when response categories follow a natural order is a complete reversal of the sequence. Unfortunately, even reversing the response sequence in one-half of the questionnaires—an approach known as the *split-ballot technique*—may not effectively neutralize potential response position bias.[6]

Response category rotation is also somewhat more difficult in a mail survey than in face-to-face or telephone interviews. When an interviewer is asking a multiple-category question, he or she can simply be instructed to start with a different response category in each interview and follow the listed sequence until the starting category is reached. For instance, in the question dealing with General Motors, the interviewer can be instructed to start with Buick and end with Oldsmobile in one interview, start with Cadillac and end with Buick in the next interview, and so on.[7] Response category rotation in mail questionnaires will require printing and mailing different versions of the same questionnaire, which may not always be feasible owing to the additional expenses involved.

Clearly, even as subtle a feature as the sequence in which response categories are presented can result in biased data. Nevertheless, depending on the nature of the response categories and the type of questionnaire administration method to be used, one may be able to reduce potential biases. Certainly a prudent procedure is to take as many precautions as feasible.

Response Category Content. What should be included in the response choices to a multiple-category question? The nature of the variable the question is attempting to investigate, as well as the type of data desired, will have a bearing on category content. However, as a general rule, the response choices must be *collectively exhaustive* (i.e., taken together, they should provide for every possible answer a respondent might give) and *mutually exclusive* (i.e., they should not overlap).

---

[5] The appropriateness of this assumption has been challenged, however. For further details, see Niels J. Blunch, "Position Bias in Multiple-Choice Questions," *Journal of Marketing Research*, 21 (May 1984), pp. 216–220.

[6] For instance, a study by William B. Locander and John P. Burton, "The Effect of Question Form on Gathering Income Data by Telephone," *Journal of Marketing Research*, 13 (May 1976), pp. 189–192, suggests that one type of ordering of numerical-response categories to a question might lead to more valid data than a reverse ordering of the same categories.

[7] When questionnaires are administered through CRT terminals (as we saw in Chapter 6), one may be able to program the central computer to rotate the response categories automatically from interview to interview; see, for example, James F. O'Hara, "CRT Interviewing: People Still Make the Difference," *The Marketing Researcher* (Walker Research), Winter 1982, p. 2.

For questions dealing with a respondent's attributes (e.g., age) or behavior (e.g., number of long-distance calls made per week), one can easily construct response categories that are collectively exhaustive and mutually exclusive. All one needs to do is to determine the total range of possible answers (which will usually be in the form of numbers) and divide the range into as many suitable categories as the researcher's data requirements dictate (the number of categories is discussed in the next subsection). A case in point is the question on long-distance telephone calls presented earlier.

Not quite so easy is constructing structured-response categories for questions dealing with respondents' attitudes, beliefs, opinions, and motives (i.e., *reasons* for behavior). The range of possible answers to such questions will usually not be clear-cut. Wording the categories in such a way as to reflect potential responses accurately may also be difficult. For example, questions like "What do you think of our president's foreign policy?" or "Why do consumers fail to use manufacturer's coupons when buying products?" will be more difficult to cast into a structured format than a question like "How many years of formal education have you completed?"

Converting open-ended questions about attitudes, beliefs, and the like into structured questions may require some exploratory research just to get a feeling for the range and the content of responses.[8] Even if such preliminary research is conducted, however, one should usually include an "other" category, just in case the structured responses listed are not collectively exhaustive. Inclusion of an "other" category, like the one in the question dealing with automatic-teller machines, will make a question partially nonstructured. But using a partially nonstructured question is better than using a fully structured question when one is not certain that the response choices provided are collectively exhaustive. Otherwise, a respondent whose intended answer does not fit into any of the listed alternatives will be forced to give an invalid response or no response at all. In either case the quality of the data collected will be adversely affected.

Another potential reason that response choices to a multiple-category question may not be collectively exhaustive is failure to include a "don't know" or "no opinion" category. When there is a possibility that some respondents may have no basis for answering a question, failure to provide a "don't know" or "no opinion" response choice will most likely lower data accuracy. As evidence for this statement, in a study wherein researchers sought respondents' opinions about a fictitious public agency, many respondents actually expressed opinions about the nonexistent organization. This finding led the researchers to conclude: "Clearly, respondents will express an opinion about issues of which they have no knowledge. Such a response occurred even in a situation involving relatively limited pressure to respond."[9] These researchers also found that inclusion of a "don't know" category reduced the occurrence of uninformed, and hence invalid, responses. Look again at the question dealing with General Motors' divisions and decide for yourself whether a "don't know" response choice should be added. Also, reexamine the other two examples of multiple-category questions. Would you recommend any modifications in them?

---

[8] This point is discussed further in A. N. Oppenheim, *Questionnaire Design and Attitude Measurement* (New York: Basic Books, 1966), pp. 25–30; see also Donald P. Warwick and Charles A. Liniger, *The Sample Survey: Theory and Practice* (New York: McGraw-Hill, 1975), pp. 136–137.

In addition to being collectively exhaustive, response choices must be mutually exclusive. That is, a respondent's answer must fit one, and only one, response category. In the following question, for instance, the response choices are collectively exhaustive but not mutually exclusive.

On the average, how many cans of beer do you drink per week?

______ 0 to 3 cans

______ 3 to 6 cans

______ More than 6 cans

Changing the 3 in the second category to a 4 will make the response choices mutually exclusive.

Now refer to the question "What do you like about First City Bank's automatic-teller machines?" and its response choices presented earlier. The meanings of the response choices are distinct and nonoverlapping. However, they overlap in another sense: Respondents' answers to the question may fall into more than one of the listed alternatives. From the way the question is currently set up, whether the researcher expects each respondent to check only one category is not clear. Consequently, while some respondents will just check one response, others will check several, even if *all* of them like more than one attribute of First City's automatic-teller machines.

Two options are available to ensure that the question is interpreted in the same way by all respondents:

1. Modify the question to read, "Which one of the following do you like *most* about First City Bank's automatic-teller machines?" With this revision each respondent would, or should, check only one response category.
2. Augment the question to read, "What do you like about First City Bank's automatic-teller machines? (Please check as many categories

---

[9] Del I. Hawkins and Kenneth A. Coney, "Uninformed Response Error in Survey Research," *Journal of Marketing Research*, 18 (August 1981), p. 373.

as apply.)" This revision explicitly indicates that multiple answers are allowed.

A researcher choosing the second option must realize, however, that coding and analyzing multiple responses may be somewhat cumbersome.[10] Which of the two revisions should be made depends on the nature of the researcher's data requirements.

Number of Response Categories. Multiple-category questions fall between dichotomous questions and open-ended questions in terms of the variety of responses they permit. Consider a question dealing with income, for instance. If the researcher just wants to know whether a respondent's income is above a certain amount, a dichotomous question will suffice. At the other extreme, if a respondent's exact income is to be ascertained, an open-ended question will be necessary. Frequently, however, a project's data requirements will be between the two extremes, and the researcher has to decide on the number of response categories to provide. Although no rules exist for determining the optimum number of categories, the following considerations are worth taking into account.

One useful approach for deciding on the number, as well as content, of response categories is to examine past studies similar to the one being contemplated. This approach can be especially beneficial if a purpose of the research is to compare the data to be generated with available secondary data. To illustrate, suppose a study of videocassette recorder owners is to be conducted to see how their demographic characteristics compare with those of the general public. Data on the general public's demographic characteristics are readily available through publications of the U.S. Census Bureau. Therefore a beneficial procedure is to make the form of the study questions (including the number of response choices) consistent with the census classifications.

Another important consideration is the type of questionnaire administration method to be used. Having too many response choices to each of numerous multiple-category questions can cause problems in a mail questionnaire. The increased questionnaire length may not only increase mailing costs but also discourage potential respondents from filling out and returning the questionnaires. The best solution under these circumstances is to list only the most likely responses (perhaps determined through exploratory research) along with an "other" category.

Multiple-category questions may pose a somewhat different set of problems in personal and telephone interviews, especially when the categories have to be read out loud to the respondents. One obvious drawback of having too many response choices when the interviewer has to read them aloud is increased interview time, which, in turn, may irritate the respondent. Another drawback is the very real possibility that the respondent may not remember all the alternatives by the time the interviewer gets to the end of the list. As a result, the answer may be biased in favor of alternatives mentioned at the very beginning or at the very end. One option for tackling these problems is to not read response categories at all—that is, to ask an open-ended question but have a collectively exhaustive list of response choices printed on the questionnaire so that coding will be easy and fast.

---

[10] This drawback is mitigated somewhat by the fact that several statistical software packages offer special coding instructions and analysis options for handling questions with multiple responses. For instance, the well-known SPSS-X package has a "MULT RESPONSE" routine for analyzing data generated by multiple-response questions. Researchers can and do use such a routine to code and tabulate data obtained from multiple-category questions that permit more than one response to be checked.

Unfortunately, the option of not reading aloud the response choices may not always be feasible. For example, in studies dealing with advertisements researchers may wish to measure *aided recall* of ads by respondents. Aided recall requires that all alternatives (e.g., all ads appearing in a certain issue of a magazine) be mentioned to the respondent. Aided recall can be accomplished adequately and efficiently in a personal interview by, for instance, handing the respondent a card listing all the alternatives and asking him or her to respond after looking through the list. This technique will obviously not be possible in a telephone interview. Thus the nature of the questions to be asked may dictate the type of questionnaire administration method to be used. For instance, a complex magazine readership and advertising recall study may leave the researcher no choice other than personal interviewing, which, as we saw in Chapter 6, is the most versatile questionnaire administration method.

QUESTION RELEVANCE AND WORDING

One of the most critical tasks after drawing up a rough questionnaire draft is to ensure the relevance of every question on it. Each question must be carefully examined to see whether the data to be generated will be pertinent to the purpose of the research. While this piece of advice seems obvious (after all, who would want to ask a question if it is irrelevant?), it is frequently overlooked, especially by inexperienced questionnaire designers. Sudman and Bradburn describe this problem:

> The process of writing questions is fun and quickly engages the interest of the participants. Competition develops among the question writers to see who can come up with the cleverest and most interesting questions. A game of "Wouldn't it be nice to know . . . ?" emerges, and soon there are many more questions than the budget can afford or the respondents' patience will endure. Too often questionnaire writers are so caught up in the excitement of question writing that they jump rapidly into writing questions before they have adequately formulated the goals of the research and thoroughly understood the research questions.[11]

---

[11] Sudman and Bradburn, *Asking Questions*, p. 13.

A worthwhile issue to raise concerning each question on the rough draft is: "Can the research objective(s) be fulfilled *without* asking this question?" If the answer is yes, the question should be deleted. The only exceptions are general open ended questions that may be asked at the start of an interview to kindle the respondents' interest or put the respondent at ease or asked during the interview to aid transition from one topic to another. If the answer is anything other than a yes (e.g., no, maybe, don't know), the question may *still* be a candidate for revision or deletion and must be further examined with respect to criteria like the ones discussed next.

Can the Respondent Answer the Question?

Even when a question is apparently relevant, it will be useless, if respondents are unlikely to have a meaningful basis for answering it. For instance, consider the following question taken from a mail questionnaire the author received from someone conducting a study of professors' perceptions about marketing education and marketing students:

> In your opinion, how many students in your marketing courses are potentially successful marketing managers?
>
> Ten percent ________
>
> Twenty percent ________
>
> Thirty percent ________
>
> ________ percent

This question deals with professors' perceptions of their students and is therefore relevant for the study. But many respondents, like the author, will be at a loss to answer it meaningfully. In other words, the data sought is likely to be beyond the realm of experience of many respondents. By the way, see if this question has any other drawbacks.

Another frequent reason for respondents' inability to answer a question, even when the information it seeks is within their realm of experience, is lack of memory. Most people can readily recall the number of traffic accidents they were involved in during the past year or even the past several years. However, most people will have difficulty remembering the number of times they have been to a McDonald's within the past year—within the previous week perhaps, but not within the past year. Ability to remember a past event depends on the time elapsed since its occurrence as well as the importance of the event to the respondent. The latter factor is crucial, because what may be extremely important to the researcher may be trivial to the respondent.

For instance, consider the following question taken from a questionnaire used by a marketing research firm in conducting a face-to-face survey of female household heads:

> How much has your family spent in the last 12 months on cookware? $ ________

No doubt this question was important to the research firm and its client. But how many respondents would have been able to remember the amount they spent on cookware during the previous 12 months? My guess is very few, if any. Yet many of them would have come up with *some* amount, whose accuracy, of course, would have been highly questionable.

The point is that good question writers, rather than become blinded by what is important to the researcher or research user, must put themselves in the respondent's shoes when evaluating a question's potential value. Questions that seem meaningless or difficult from the respondent's standpoint must be modified or deleted.

Will the Respondent Answer the Question?

When a question deals with a sensitive issue or is embarrassing, respondents may refuse to answer it, even if they have sufficient information to respond accurately. Worse still, they may actually provide less-than-truthful responses. Questions about personal financial matters or sexual behavior are examples of sensitive questions. Such questions also run the risk of irritating the respondents. Therefore, unless they are absolutely essential, they should not be included in the questionnaire. In studies where data on sensitive issues are crucial, the researcher may have to use disguised techniques (such as the ones discussed in Chapter 6) or some other special types of questioning.[12]

The relevance of a question, the respondents' ability to answer it, and their willingness to answer it can all be affected by its *wording*. Writing questions capable of generating relevant and valid data is perhaps the most critical and most difficult part of designing a questionnaire. Unfortunately, there is no magic wand one can wave to produce the *best* question(s) to fit each research objective. Most books on questionnaire design devote numerous pages to the dos and don's of proper wording.[13] Yet as Dillman observes:

> Writing questions for a particular questionnaire means doing them for (1) a particular population, (2) a particular purpose, and (3) placement next to other questions in the questionnaire. Words that are too difficult for use with some populations may be perfectly acceptable for others. . . . A question that makes little sense by itself may be quite clear when asked after the ones which precede it in the questionnaire. . . . A list of admon-

---

[12] Such special types of questioning are not needed frequently in marketing research studies. A good source of further information is Sudman and Bradburn, *Asking Questions*, Chap. 3; see also Ed Blair, Seymour Sudman, Norman M. Bradburn, and Carol Stocking, "How to Ask Questions About Drinking and Sex," *Journal of Marketing Research*, 14 (August 1977), pp. 316–321; James E. Reinmuth and Michael D. Geurts, "The Collection of Sensitive Information Using a Two-Stage Randomized Response Model," *Journal of Marketing Research*, 12 (November 1975), pp 402–407.

[13] For example, see Payne, *Art of Asking Questions;* Sudman and Bradburn, *Asking Questions*.

itions, no matter how well intended, therefore cannot be considered as absolute principles that must be adhered to without exception.[14]

In short, the wording of a question and its influence on responses may be so situation-specific and complex as to defy any universal set of guidelines. For this reason, and also because of space limitations, we will not cover all the issues of question wording here; books dealing specifically with questionnaire design offer excellent discussions. Rather, in the following sections we will examine several examples to highlight some of the most frequent errors stemming from question wording and the ways such errors can be minimized.

Double-Barreled Questions

Suppose the following question is to be used in conducting a survey of the general public:

> Do you feel firms today are concerned about their employees and customers?
>
> ________ Yes ________ No ________ No opinion

Double-Barreled Question *A question that raises several separate issues but only provides one set of responses.*

This question is a classic example of a ***double-barreled question***, one that raises several *separate* issues but only provides one set of responses.

Answers to double-barreled questions are very difficult to interpret. For instance, suppose a respondent answers no to this question. A "no" response can be interpreted three different ways:

- The respondent feels firms are concerned about neither employees nor customers.
- The respondent feels firms are concerned about employees but not customers.
- The respondent feels firms are not concerned about employees although they are concerned about customers.

Which of these interpretations is accurate? Only the respondent will know for sure. A "yes" or a "no opinion" response to this question may be equally difficult to interpret.

If meaningful, easy-to-interpret data are to be obtained, a double-barreled question must be reworded to focus on one specific issue at a time and must provide an unambiguous set of response choices. This revision is usually accomplished by breaking it up into several questions. Thus our illustrative question can be reworded as two separate questions:

---

[14] Dillman, *Mail and Telephone Surveys,* pp. 96–97, excerpted with permission from John Wiley & Sons, Inc. Several controlled studies have found that the *context* in which a question is asked will affect the nature of the response; see, for example, Howard Schuman, Stanley Presser, and Jacob Ludwig, "Context Effects on Survey Responses to Questions About Abortion," *Pubic Opinion Quarterly,* 45 (Summer 1981), pp. 216–223.

1. Do you feel firms today are concerned about their employees?

__________ Yes __________ No __________ No opinion

2. Do you feel firms today are concerned about their customers?

__________ Yes __________ No __________ No opinion

Leading Questions

Leading (Loaded) Question
A question that tends to steer respondents toward a certain answer, irrespective of what their true answers are.

A ***leading question***, also known as a ***loaded question***, is one that may steer respondents toward a certain answer, irrespective of what their true answers are. The following questions illustrate:

Don't you think offshore drilling for oil is environmentally unsound?

__________ Yes __________ No __________ No opinion

Do you think the quality of products on the market today is as high as it used to be ten years ago?

__________ Yes __________ No __________ No opinion

Of course, not all respondents may be led to a certain answer by questions like these. Respondents who have very definite feelings about the environmental impact of offshore drilling or about the quality of products currently on the market are unlikely to be swayed by the way the questions are worded. However, other respondents including those who really have no opinion, are likely to be tempted to answer yes to the first question and no to the second. Invalid data will result if those respondents yield to the temptation.

The risk of asking leading questions is especially great when the focus of research is on people's attitudes, beliefs, or opinions. A researcher may end up designing leading questions often unknowingly, by failing to maintain an objective posture during the questionnaire design process—that is, by wording questions in such a way as to bias potential responses in favor of his or her own preconceived notions about the issues being studied. A conscious effort to construct questions in as neutral a fashion as possible is a prerequisite for obtaining unbiased data.

The two leading questions we saw earlier can be modified for neutrality as follows:

What is your feeling about the environmental impact of offshore drilling for oil?

__________ Offshore drilling is environmentally sound.

__________ Offshore drilling is environmentally unsound.

__________ No opinion.

How do you feel about the quality of products on the market today compared with the quality of products on the market ten years ago?

________ Product quality is better now than ten years ago.

________ Product quality is worse now than ten years ago.

________ Product quality is the same now as ten years ago.

________ No opinion.

One-Sided Questions

One-Sided Question *A form of leading question that presents only one aspect of an issue for which respondents' reactions are being sought.*

As their name implies, ***one-sided questions*** present only one aspect of an issue for which respondents' reactions are being sought. One-sidedness can bias the response, often very subtly. Consider the following questions (notice that the first question is one of the two we came up with in modifying the double-barreled question presented earlier):

Do you feel firms today are concerned about their employees?

________ Yes ________ No ________ No opinion

Would you agree or disagree that deregulation of the airline industry has benefited consumers?

________ Agree ________ Disagree

________ No opinion

To appreciate why these questions are one-sided, and hence may lead to biased responses, look at the following rewordings of the same questions:

Do you feel that firms today are concerned about their employees or that they are unconcerned about their employees?

________ Concerned ________ Unconcerned

________ No opinion

Do you feel deregulation of the airline industry has benefited consumers, has had no impact on consumers, or has hurt consumers?

________ Has benefited consumers.

________ Has had no impact on consumers.

________ Has hurt consumers.

________ No opinion.

The reworded questions are more neutral in that they make explicit all sides of the issues involved. Intuitively, they should be better than their one-sided counterparts, which may nudge the respondents toward the side presented. One-sided questions are, in a way, leading questions, although they may lead respondents in a more subtle fashion.

Several studies have shown that respondents, especially those with low levels of education, have a tendency to agree with whatever side is presented by one-sided questions.[15] This phenomenon is referred to as ***yea-saying***, and the resulting bias is known as ***acquiescence bias***. The rationale behind presenting all sides of an issue is that the respondent will at least have something to think about instead of blindly answering yes or agreeing with whatever the question says. Presenting all sides of an issue should reduce erroneous responses.

Acquiescence Bias (Yea-Saying) *A bias stemming from the tendency on the part of certain respondents to agree with whatever side is presented by one-sided questions.*

Split-Ballot Technique *A procedure wherein (1) the study sample is split into two equivalent halves and (2) a different version of the same questionnaire (e.g., different sequencing of questions) is administered to each half.*

The ***split-ballot technique***, mentioned in our earlier discussion about rotating response choices in a multiple-category question, has also been suggested as a means of neutralizing possible acquiescence bias.[16] In using this technique to nullify acquiescence bias, two versions of the questionnaire are prepared: one version with questions presenting one side of the issues being researched and the second version with questions presenting the other side. (While the split-ballot technique usually involves only two versions of the same questionnaire, as many versions as there are alternatives to be presented can be created.) The respondent sample is also randomly split into two comparable halves, and each half is administered one version of the questionnaire. The data obtained from the two halves are merged with the hope that any acquiescence biases will cancel out.

Although the split-ballot technique appears elegant, it has a couple of limitations. First, preparing different versions of the questionnaire can become complicated and quite expensive if several different versions are needed. Second, if responses to the same question vary widely across the different versions, combining the responses to yield an average response may not be meaningful, in which case the question will be essentially useless. Therefore the split-ballot technique is not recommended, unless stating all sides of an issue within the same question is likely to make the question too complex or awkward.

A variation of one-sidedness can occur in multiple-category questions when the alternatives presented are loaded toward one side, as in the following example:

---

[15] See, for example, George F. Bishop, Robert W. Aldendick, and Alfred J. Tuchfarber, "Effects of Presenting One Versus Two Sides of an Issue in Survey Questions," *Public Opinion Quarterly*, 46 (Spring 1982), pp. 66–85; Alfred M. Falthzik and Marvin A. Jolson, "Statement Polarity in Attitude Studies," *Journal of Marketing Research*, 11 (February 1974), pp. 102–105; Warwick and Liniger, *The Sample Survey*, p. 146.

[16] See, for example, Falthzik and Jolson, "Statement Polarity in Attitude Studies."

How important is price to you in buying a new car?

________ More important than any other factor.

________ Extremely important.

________ Important.

________ Somewhat important.

________ Unimportant.

The alternatives listed are *unbalanced*: Four out of the five suggest price is an important criterion.[17]

To understand the potential bias that unbalanced alternatives may subtly introduce, consider the following version of the same question:

How important is price to you in buying a new car?

________ Very important.

________ Relatively important.

________ Neither important nor unimportant.

________ Relatively unimportant.

________ Very unimportant.

The balanced alternatives in this version offer a wider choice to respondents who do not consider price to be a critical factor. Stated differently, respondents who feel that price is relatively unimportant or is neither important nor unimportant, may be *forced* to check "somewhat important" in the unbalanced version of the question for lack of a more suitable category. This response can lead to erroneous interpretations: The researcher has no way of knowing whether respondents who indicated "somewhat important" really meant to give that response. Clearly, a prudent procedure is to provide respondents with as balanced a set of alternatives as possible—and thereby reduce the chances of drawing invalid inferences from the data collected.

Questions with Implicit Assumptions

The responses to some questions can be greatly influenced by what respondents assume in answering them. ***Questions with implicit assumptions*** do not provide, or imply, the same frame of reference to all respondents. Thus the question should state explicitly what respondents should assume, rather than let them make their own assumptions. Suppose the following question is part of a questionnaire to be mailed to a city's residents:

Question with Implicit Assumptions *A question that does not provide or imply the same frame of reference to all respondents.*

---

[17] Payne, *The Art of Asking Questions,* pp. 90–92, contains a more detailed discussion and additional examples of questions with unbalanced alternatives.

Are you favorable, indifferent, or unfavorable toward a 10% increase in city taxes?

________ Favorable ________ Indifferent

________ Unfavorable

A respondent's answer to this question will differ depending on whether he or she has accurate information, inaccurate information, or no information about why the 10% tax increase is needed. The researcher must provide a common frame of reference to all respondents in order to obtain meaningful responses and draw valid inferences from them: To assume that all respondents have the same accurate background information pertaining to the question is dangerous. The question on city taxes can be improved as follows:

Are you favorable, indifferent, or unfavorable toward a 10% increase in city taxes to repair potholes in the city's streets?

________ Favorable ________ Indifferent

________ Unfavorable

Implicit assumptions may be embedded in a question in more ways than one. Consider the following question:

Do you drink beer while watching football, baseball, and so on?

________ Yes ________ No

This question makes several implicit assumptions. First, from the respondent's standpoint it is not clear whether "watching football, baseball" applies to doing so at home (i.e., on TV), outside the home, or both. Therefore the answer is likely to vary depending on what a respondent assumes. Second, what exactly "and so on" includes is unclear. Does it, for example, include or exclude sporting events other than ball games? Again, the question is subject to varying interpretations by different respondents. Third, the question apparently assumes that the respondent *is* a beer drinker. Whether this implicit assumption is a problem depends on what questions were asked before this particular one. The first two problems, concerning vagueness of the question, may be remedied by rewording the question as follows:

During which of the following activities do you drink beer? (Check as many as apply.)

________ Attending ball games.

________ Attending sporting events other than ball games.

________ Watching ball games on TV.

________ Watching sporting events other than ball games on TV.

The number and types of response categories included in this question depend on the researcher's specific objectives. But the important point is that the reworded format has much less room for varying assumptions across respondents.

Filter Question *A question meant to qualify respondents for a subsequent question or to ensure that the question is within their realm of experience.*

Let us now turn to the apparent assumption that the respondent is a beer drinker. The question will obviously be irrelevant to respondents who do not drink beer at all. One way to avoid asking this question of such respondents is to use a filter question. A ***filter question*** is meant to qualify respondents for a subsequent question or to ensure that the question is within their realm of experience. A possible filter question in our example is: "Do you ever drink beer?" Only those who answer yes will be asked the question about activities during which they drink beer.

One note of caution concerning filter questions is that too many of them can sharply increase questionnaire length and interview time. They should therefore be used sparingly—perhaps only when they are needed to qualify respondents for the entire interview or to avoid asking a question of an ineligible respondent that can be ridiculous or embarrassing (e.g., asking a respondent how many children he or she has before ascertaining whether he or she has ever been married). Under other circumstances any potential problem of lack of question relevance can be overcome by simply adding a special response category (in lieu of a separate filter question) to be checked by respondents if the question does not apply. For instance, in the multiple-category version of the question on beer drinking, the following response choice can be provided at the *top* of the list:

________ I do not drink beer. (Go on to the next question.)

The following question is yet another illustration of a question that is vague as a result of the question writer's failure to make explicit what exactly certain key words include or refer to:

How often do you eat eggs for breakfast?

________ Frequently

________ Occasionally

________ Rarely

________ Never

The problem with this question is that different respondents may attach different meanings to the first three response choices. For instance, one respondent's notion of "occasionally" may be what another respondent may term "rarely." When such varying interpretations are possible, the meaningfulness of the data collected will be questionable. One way of guarding against errors due to misinterpretations is to revise the question as follows:

On the average, how many days per week do you eat eggs for breakfast?

_________ Every day.

_________ 5 or 6 days.

_________ 3 or 4 days.

_________ 1 or 2 days.

_________ Less than 1 day per week.

_________ Never eat eggs for breakfast.

Of course, the question writer may not always be able to come up with categories that are as clearly spelled out as in this instance. But the point is that to the extent possible the question writer must choose words and categories that will be interpreted the same way by all respondents.

Complex Questions

Words that a question writer may understand perfectly may be unfamiliar or sound complicated to the respondent. Using such words in a question will make it ***complex***, as the following example shows:

Complex Question
A question that is difficult to answer because it contains terms unfamiliar to respondents and/or requires considerable respondent effort.

In which of the following do you typically invest your liquid assets?

_________ Insured accounts

_________ Stock market

_________ Insured accounts and stock market

_________ Other accounts

If this question is posed to respondents drawn from the general public, many of them may not know what "liquid assets" mean, and some of them may even be unfamiliar with terms like "stock market." Nevertheless, most of them will probably give some response so as not to reveal their ignorance. One obviously cannot have much trust in data obtained from respondents who answer a question without understanding it. The moral is, Use simple words, when possible, in writing questions and make sure respondents can easily understand them. However, this guideline cannot be used rigidly. As Dillman correctly points out: "Substituting simple for complex words often has the paradoxical effect of turning simple sentences into complex ones."[18]

In trying to cast a question in the simplest possible terms, the writer must also ensure that it does not get complicated because of increased length. Stated differently, in the interest of keeping questions relatively short, one may use complex words as long as potential respondents are likely to be familiar with them. A sample of business executives, for example, should have no difficulty understanding the terms *liquid assets* and *stock market*. In fact, when the respondents are knowledgeable, spelling out such terms in simple words may be viewed as an insult.

---

[18] Dillman, *Mail and Telephone Surveys*, p. 97.

In short, whether a question is perceived as simple or complex depends not only on its wording but also on the type of respondent sample to which it will be posed. A good question writer must be sensitive to respondents' capabilities and design questions that neither go over their heads nor talk down to them.

Another type of complex question is one that demands too much effort from the respondents. An example of such a question is as follows:

> Of the total number of miles you drove during the past month, approximately what percentage was for driving to and from work?
>
> __________ percent

This question is complex not because it is difficult to understand but because it is hard to answer. A respondent has to expend a lot of mental effort to answer it. An approach to simplifying it somewhat is to break it into two questions: one focusing on the total number of miles driven and the other focusing on the number of miles driven to and from work. The desired percentage figure can then be computed by the researcher during the data analysis stage. Whenever feasible, one should ask several simple quetions rather than one complex question. The less work the respondents have to do, the greater is their willingness to answer a question and the lower is their chance of making errors in responding.

In summary, despite our lengthy treatment of question wording, we have merely scratched the surface of this topic. However, our discussion of the various types of faulty questions—double-barreled questions, leading questions, one-sided questions, questions with implicit assumptions, and complex questions—does illustrate the most common wording problems and highlights the dangers of not remedying them. Some of these same problems, or variations of them, are bound to be present in most rough questionnaire drafts. Research In Use 10.1 provides several illustrations of wording problems and what researchers are doing to minimize them.

SEQUENCING OF QUESTIONS

Questions must be arranged in a logical sequence in order to minimize data errors and to facilitate easy and smooth administration of the questionnaire. The ordering of specific questions in a given situation, of course, depends on characteristics unique to that situation. However, the broad guidelines presented in this section should be helpful in virtually all situations.

Research in Use 10.1 *Accuracy of Survey Data*

"Excuse Me, What's the Pollster's Big Problem?"

You're cornered in the street by a pollster who hands you a photo of Ronald Reagan standing next to David A. Stockman. "You know who this man is, don't you?" he demands. If you are like most people, your answer is prompt: "The President of the United States." But what if he asked: "Do you have any idea at all who this man is?" When Stanford University psychologist Herbert H. Clark tried that on 15 volunteers, only one identified Reagan. Most identified Stockman—or asked for more information.

Polls and market surveys are a black art. Ask the questions one way, and the answers reflect what the public is thinking. But phrase them just slightly differently, and the results can be totally out of touch with reality. In 1984 a panel of experts from the National Research Council concluded that survey designers "simply do not know much about how respondents answer questions." Adds sociologist Judith Tanur of the State University of New York at Stony Brook: "Survey designers have taken a seat-of-the-pants approach to constructing questionnaires. It hasn't been guided by a framework or a theory."

Bad Memory. But just as the ancient Greeks consulted oracles, companies and government agencies are more frequently turning to professional pollsters to augur everything from election results to product introductions. "Business is definitely up," says Harry W. O'Neill, vice-chairman of Opinion Research Corp. in Princeton, N.J. Last year the top 44 U.S. polling companies pulled in more than $1.5 billion in revenues. While revenues at most polling companies have been growing by at least 8% annually since 1983, some pollsters, such as Gallup Organization Inc. in Princeton, report gains of 15% to 20%.

Policy decisions and billions of dollars in corporate funds ride on the accuracy of those polls. So the people who design them are teaming up with researchers in cognitive psychology, anthropology, linguistics, and decision theory. At least a dozen research projects are under way at universities such as Yale, Stanford, and the University of Chicago.

Already, some of the major pollsters as well as several federal agencies, including the Bureau of Labor Statistics and the Census Bureau, are using the new techniques to design better surveys. Gallup President Andrew Kohut says his company has just made a "major investment" in a company that builds devices for measuring eye movement. Using it, he says, Gallup researchers hope to study the ways people respond to television commercials and magazine ads.

The researchers have discovered that it's a big mistake to rely on people's ability to remember events. Stanford psychologist Lee D. Ross discovered that students were only marginally better at recalling details, such

(Continued)

Research in Use 10.1 *Continued*

as how many checks they'd written during the course of a month, than they were at predicting the future behavior of their roommates. In addition, the mind tends to "telescope" past events: They often seem to be more recent than they really are. "Most of us just don't have a good internal calendar for dating events," observes David J. Mingay, a cognitive psychologist at the National Center for Health Statistics in Hyattsville, Md.

Mingay is part of a team that is designing better questions for an annual health survey that polls 135,000 Americans. He is exploring a technique that measures the time lag between when a question is asked and when it is answered. Studies show that a too-slow response may mean the question is too difficult to understand, while an immediate response probably means that it was either misunderstood or hurriedly answered. Pollsters are also turning to a technique known as "benchmarking" to help subjects place events in time. For example, surveyors found that answers to questions such as "How many times have you been robbed since the Chernobyl disaster?" are more accurate than answers to "How many times have you been robbed within the past 12 months?"

Just a Guess. The survey designers are also trying to unravel the thought processes that lead to an answer. By taping test subjects' answers and then asking them to explain why they responded in a particular way, researchers can determine whether a question is eliciting the information that pollsters are trying to examine. Others ask subjects to rate answers according to how confidently they feel about them. A low score usually means the answer is just a guess.

Pollsters have an added incentive for honing their questions. The public is so barraged by surveyors that people refuse to be guinea pigs about 40% of the time. Improving accuracy, however, is still the most important goal of pollsters. This year the BLS will spend about $200 million on more than 50 surveys, which Commissioner Janet L. Norwood says will be "the most politically and economically sensitive surveys in the U.S. government." And with so much riding on the results, she adds, "we can't afford to be wrong."

Source: Article by Sana Siwolop. Reprinted from February 16, 1987 issue of *Business Week* by special permission, Copyright © 1987 by McGraw-Hill, Inc.

Position of Demographic and Sensitive Questions

Classification Data *Data generated by questions dealing with respondents' personal or demographic characteristics.*

Place questions about respondents' personal or demographic characteristics (e.g., age, education level, and income) at the end of a questionnaire. These questions are included in virtually all questionnaires since they provide ***classification data*** useful in obtaining a profile of the respondent sample and in cross-classifying responses to other questions that pertain more directly to the study objectives. However, asking these questions at the beginning may irritate some respondents and affect their willingness to complete the rest of the survey. The only situation in which demographic questions should be asked at the beginning is when they are to serve as filter questions to qualify respondents for the survey. This situation will arise when respondents are chosen through quota sampling (discussed in Chapter 14).

Sensitive questions likely to embarrass respondents or put them in an awkward position should also be placed near the end of the questionnaire. This placement is especially critical if the questionnaires are to be administered through face-to-face or telephone interviews. Good rapport between interviewer and respondent is essential for obtaining truthful responses to sensitive or threatening questions. So a delay in asking sensitive questions will offer the interviewer additional time to build rapport with the respondent. Placing sensitive questions toward the end of a questionnaire has another advantage: Even if the respondent refuses to answer such questions, data gathered through earlier questions may still be useful. But asking sensitive questions at the beginning may cause the entire interview to be lost.

Ask simple questions as early as possible. This guideline is a corollary of sorts to the preceding guidelines. Easy-to-answer quetions can get the questionnaire administration process off to a smooth start (even in the case of mail questionnaires, which are self-administered). They also help pave the way for more difficult questions.

Arrangement of Related Questions

When a questionnaire addresses a variety of topics, clustering questions that focus on the same topic is advisable. Skipping from topic to topic in a haphazard fashion may confuse respondents, break their train of thought, and cause errors in the data. In other words, grouping questions into meaningful clusters can increase respondents' ease in answering the questions and reduce the chance of response errors. For instance, in a survey of business executives' views on the economy, competition, and employee turnover, a prudent approach is to divide the questionnaire into three parts, with questions in each linked to a single topic.

Funnel and Inverted-Funnel Sequences

Funnel Sequence
An ordering of questions that begins with a very general question on a topic and gradually leads up to a narrowly focused question on the same topic.

Move from general to specific questions within a topic. This approach is also known as using a ***funnel sequence***.[19] It involves beginning with a very general question on a topic and gradually leading up to a narrowly focused question on the same topic. Employing a funnel sequence is advisable in virtually all situations (with one exception, which we will consider a little later). But it is *essential* in situations where asking specific questions first can bias the answers to later questions.

---

[19] Sudman and Bradburn, *Asking Questions,* pp. 219–221, contains a good discussion of the funnel sequence and its pros and cons.

To illustrate, consider the following sequence of multiple-category questions. (*Note:* Only illustrative answer categories have been shown for each question to conserve space.)

1. Which of the following types of TV shows do you watch? (Check as many categories as apply.)

_________ News shows _________ Quiz shows

_________ Game shows

2. Which of the following types of TV shows do you *like the most?* (Check as many categories as apply.)

_________ News shows _________ Quiz shows

_________ Game shows

3. Which of the following specific shows did you watch during the past seven days? (Check "yes" or "no" for each show listed.)

"All in the Family" _________ Yes _________ No

"The Price Is Right" _________ Yes _________ No

"CBS Evening News" _________ Yes _________ No

These three questions follow a funnel sequence. Reversing this sequence (i.e., asking first about the specific shows watched during the past seven days) may bias responses to the questions on the *types* of shows watched and liked most. That is, respondents may intentionally or unintentionally give more weight to the types of shows similar to the specific ones they recalled watching during the past seven days.[20]

Inverted-Funnel Sequence
An ordering of questions that begins with very specific questions on a topic and moves on to more general questions on the same topic.

Although a funnel sequence is the generally recommended format for ordering questions pertaining to the same topic, it may not always be appropriate. An ***inverted-funnel sequence*** may be better when respondents do not have clearly formulated views about a topic or when they need to have a common frame of reference in responding to general questions on the topic.[21]

For example, consider a survey of consumers to measure the image they have of XYZ Company, one of numerous other companies of similar size within the same industry. Given the proliferation of companies, many respondents may be at a loss to readily express what they feel about any particular company, including XYZ. Furthermore, if they are first asked a general-image question (e.g., "What is your overall impres-

---

[20] Studies have shown that responses to questions dealing with people's attitudes and opinions are especially likely to be influenced by the location of the questions within a questionnaire; see, for example, Lee Sigelman, "Question-Order Effects on Presidential Popularity," *Public Opinion Quarterly,* 45 (Summer 1981), pp. 199–207; Schuman, Presser, and Ludwig, "Context Effects on Survey Responses."

[21] See Sudman and Bradburn, *Asking Questions,* pp. 219–221.

Exhibit 10.3 *Question Sequence Needing Improvement*

11. Do you own or rent your current place of residence?

Own Rent

12. How long have you lived in this state?

_________ Less than one year

_________ One year to less than 5 years

_________ 5 years or more

13. If you have lived outside this state, in which state did you live immediately before moving here?

____

14. If you have lived in this state 5 years or more and if you currently rent your residence, do you intend to buy a home within the next two years?

_________ Yes _________ No _________ Don't know

15. How long have you lived in your current place of residence?

_________ Less than one year

_________ One year to less than 5 years

_________ 5 years or more

sion of XYZ Company?"), they may not all use the same frame of reference in answering it. For instance, some may base their responses on XYZ's ads; others may base their responses on XYZ's products. Therefore, it may be preferable to ask the respondents what they think of specific attributes of XYZ (e.g., ads, products, employees) *before* asking them for their overall opinions. This inverted-funnel sequence may help jog the respondents' memories and may provide them with a common set of criteria on which to base their overall opinions.

Skip Patterns

Ensure that question sequencing is conducive to clear and simple skip patterns. In many questionnaires not all questions may apply to all respondents. Questions that are not relevant to certain respondents must be skipped. Consequently, the number of questions asked and the sequence in which they are asked may vary across respondents and give rise to different skip patterns. Proper sequencing of questionnaire items is essential to avoid complicated skip patterns that may confuse interviewers and respondents.

Exhibit 10.4 *Improved Question Sequence*

---

11. Have you always lived in this state?

_______ Yes (Go to Question 12) _______ No (Go to Question 11a)

11a. In which state did you live immediately before moving into this state?

____

12. How long have you lived in this state?

_______ Less than one year

_______ One year to less than 5 years

_______ 5 years or more

13. Do you own or rent the place where you live?

_______ Own (Go to Question 15) _______ Rent (Go to Question 14)

14. Do you intend to buy a home within the next two years?

_______ Yes _______ No _______ Don't know

15. How long have you lived in your current place of residence?

_______ Less than one year

_______ One year to less than 5 years

_______ 5 years or more

To illustrate this guideline examine Exhibit 10.3, which contains part of a mail questionnaire. The five questions appear straightforward at first glance. Notice, however, that questions 13 and 14 require the respondent to first determine whether they should even be answered. Making respondents screen themselves for a question by using *if* statements has two undesirable features: (1) the respondents have to do extra work to decide whether or not to answer the question, and (2) too many *ifs* in the filter part of the question (e.g., as in question 14) may confuse respondents and cause errors. While question *wording* is crucial in preventing these potential problems, question *sequencing* is equally important, as discussed next.

Exhibit 10.4 presents a modified version of the partial questionnaire in Exhibit 10.3. The revised question sequencing and the corresponding wording changes have simplified the questionnaire. The *if* filters requiring a certain amount of effort from respondents to screen themselves have been replaced with simpler *go to* instructions at appropriate places. Furthermore, the question sequencing appears more logical than in the original version. Moving question 11 in the original version to the position of question 13 in the revised version offers a major benefit: Questions dealing with the same issue are now closer together. While all the questions deal with respondents' places of residence, questions 11, 11a, and 12 deal with the state of the residence, and questions 13, 14, and 15 deal with the building of the residence. The questions also follow a funnel sequence: state questions first, followed by building questions.

The preceding observations about Exhibit 10.4 have an important message to questionnaire designers: Just following the various question-wording and -sequencing guidelines outlined earlier (e.g., ask simple questions, cluster similar questions) can be very helpful in devising skip patterns that are clear and simple.

You probably have noticed that the skip patterns in Exhibits 10.3 and 10.4 are not equivalent. Specifically, in the original version only respondents who currently rent *and* have lived in the state for 5 years or more are required to answer question 14. In the revised version, however, *all* respondents who currently rent are required to answer question 14. Is this revision a problem? It is not really a problem, for three reasons. First, question 14 is appropriate and meaningful to all respondents who currently rent. Second, and perhaps more importantly, introducing an additional filter to qualify current renters on the basis of length of residence within the state will unduly complicate the questionnaire. Third, the researcher can still selectively ascertain the home-buying intentions of only those renters who have lived in the state for 5 years or more by cross-tabulating the responses to questions 12 and 14 (more on cross-tabulation in Chapter 16).

The point is that in the interest of keeping a questionnaire simple and minimizing potential confusion, filters are best used only when a subsequent question is likely to be meaningless or embarrassing to certain respondents. Alternatively, keeping the number of filter questions to a minimum, through appropriate wording and sequencing of questions, can facilitate questionnaire administration and reduce the chances of errors.

QUESTIONNAIRE APPEARANCE AND LAYOUT

How a questionnaire looks and how questions are laid out within it can influence the degree of respondent cooperation as well as the quality of the data collected. Appearance and layout are especially critical in mail surveys, because the questionnaire has to sell itself. A professionally done, attractive mail questionnaire can increase the chances of respondent cooperation. Furthermore, an uncluttered questionnaire—one with clear instructions, adequate separation between questions, properly located answer spaces, and so on—will significantly lower the chances of errors.[22]

To illustrate, consider the following two versions of the same question:

Version 1.

How old are you?

_____ Less than 18 _____ 18 to 25 _____ 26 to 40 _____ Over 40

Version 2.

How old are you?

_____ Less than 18
_____ 18 to 25
_____ 26 to 40
_____ Over 40

Version 2 is laid out better than version 1. Also, it is less likely to lead to the error of inadvertantly checking the wrong category (e.g., a respondent who is thirty years old checking the "over 40" category—a serious error, but one that cannot be detected by the researcher). Clearly, paying careful attention to questionnaire layout before data collection begins is worthwhile. Yet layout considerations are often overlooked, especially by those new to questionnaire design.

A detailed discussion of questionnaire features such as type of paper, type of printing, and paper size is not given here but can be found elsewhere.[23] In general, a questionnaire must appear attractive, neat, and uncluttered. It must also be convenient to handle, easy to read, and simple to fill out.

An important consideration in deciding how a questionnaire should look is *cost*. As you might expect, making a questionnaire look professional and eye-catching can be expensive. However, cost is usually more of a constraint on appearance (e.g., quality of paper and printing) than on layout (e.g., arrangement of questions and answer categories). Therefore a limited budget is *not* an excuse for a cluttered and confusing questionnaire.

---

[22] An interesting example of how a seemingly unimportant layout consideration led to significant response errors is described by Charles S. Mayer and Cindy Piper, "A Note on the Importance of Layout in Self-Administered Questionnaires," *Journal of Marketing Research*, 19 (August 1982), pp. 390–391.

[23] See, for example, Dillman, *Mail and Telephone Surveys*; Paul L. Erdos, *Professional Mail Surveys* (New York: McGraw-Hill, 1970).

We next turn to pretesting, which can point out potential problems with questionnaire layout and with question wording and sequencing.

PRETESTING

Pretesting *An important stage in the questionnaire refinement process, prior to finalizing the questionnaire; it involves administering the questionnaire to a limited number of potential respondents and other knowledgeable individuals in order to identify and correct design flaws.*

Pretesting involves administering a questionnaire to a limited number of potential respondents and other individuals capable of pointing out design flaws. It is indispensable because even the most diligent questionnaire designer may make mistakes that can only be detected through an external evaluation. While most researchers recognize the importance of pretesting, it is often improperly conducted, or even misused, as discussed next.

A common misuse of pretesting stems from viewing it as a *substitute* for careful thought and attention in the earlier stages of questionnaire design. This view can lead to a false sense of security about the soundness of a questionnaire. As Payne aptly puts it:

> The value of [pre]testing lies first in knowing the points for which to test. Knowledge of considerations that go into the original wording is essential to good testing. . . . Wording and testing necessarily go hand in glove. To shift all responsibility from one to the other would, like taking off the glove, let both hand and glove go cold.[24]

Pretesting must be viewed as a tool for shedding light on specific features or issues in a questionnaire that the researcher is particularly concerned about. For example, are the answer categories to question XX collectively exhaustive? Is such and such a word likely to be understood in the same way by all respondents? Are instructions to skip such and such a question clear? It must not be viewed solely as a general "fishing expedition" capable of catching whatever flaws the questionnaire may have.

To assume that pretest respondents will be able to uncover all the potential limitations of a questionnaire is incorrect. In fact, one study found that pretest respondents failed to detect even glaring errors, especially those concerning questions with loaded and ambiguous terms. According to the authors of this study: "In case of ambiguity, a respondent may not realize that more than one meaning can be associated with a particular term. Because the error arises from different meanings being used by different respondents, a single respondent would be unlikely to bring this error to the attention of the interviewer."[25] Thus all errors stemming from differences in interpretation across respondents will go undetected in a pretest unless the researcher scrutinizes the questionnaire beforehand, identifies potential trouble spots, and specifically probes the pretest respondents about those trouble spots. Stated differently, a researcher who considers a questionnaire to be sound simply because pretest respondents filled it out completely without any apparent difficulty is making a big mistake.

---

[24] Payne, *The Art of Asking Questions*, pp. 13–14.

[25] Shelby D. Hunt, Richard D. Sparkman, Jr., and James B. Wilcox, "The Pretest in Survey Research: Issues and Preliminary Findings," *Journal of Marketing Research*, 19 (May 1982), pp. 269–273; the quotation is from p. 272.

There are no standard specifications for the number and the nature of pretests to conduct in a given situation. However, generally, one pretest should be conducted by using personal interviews, irrespective of the administration methods to be used utlimately, because being face to face with respondents may suggest problem areas or points of confusion that may otherwise go unnoticed.[26] Also normally recommended is that a second pretest be conducted using the proposed questionnaire administration method. The purpose of this pretest is to detect problems that may be unique to the way in which the questionnaire is to be administered. Finally, when a questionnaire draft is substantially modified from the results of any pretest, one or more additional pretests may be necessary before the questionnaire is finalized.

How many respondents should be included in the pretest sample? Pretest sample size is a subjective decision that depends on a variety of factors, such as how confident the researcher is that the questionnaire is sound and the time and money available. In general, however, it is better to pretest the questionnaire systematically (i.e., by having specific objectives in mind and by extensive probing of respondents) on a relatively small sample than to pretest it on a relatively large sample by simply asking the respondents to fill it out. In other words, the potential usefulness of pretesting will depend more on quality than on quantity.

Regarding the composition of pretest respondents, they should be similar to respondents who will ultimately participate in the study. However, pretesting need not be limited to such respondents. Other individuals capable of providing valuable insights include the researcher's colleagues as well as potential users of the data to be gathered.[27] Pretesting the questionnaire on them can be extremely valuable since they are likely to look at the questionnaire more critically than the typical survey respondent. In fact, a worthwhile approach is to first obtain feedback from such expert respondents and modify the questionnaire draft as necessary *before* pretesting it on survey respondents.

Computerized Interviewing *A data collection procedure in which the questionnaire appears on a CRT terminal and the responses are directly entered into computer memory.*

QUESTIONNAIRES FOR COMPUTERIZED INTERVIEWING

Recent technological advances are leading to increased use of ***computerized interviewing*** in which the questionnaire appears on a CRT terminal and the responses are directly entered into computer memory. Examples of computerized interviewing include centralized telephone interviewing with the aid of CRT terminals and questionnaires filled out by respondents sitting at computer keyboards in shopping malls. Although questionnaires used in computerized interviewing are stored in computer memory, they still have to be designed first, bearing in mind the guidelines and caveats discussed thus far. Also, properly programming the questionnaire into computer memory requires additional time and expense. Nevertheless, computerized interviewing, in addition to eliminating the need for printed questionnaires and enabling instant analysis of the collected data, offers several attractive questionnaire features not possible in the past. The following list illustrates these features.

---

[26] Ironically, however, the study by Hunt, Sparkman, and Wilcox, "The Pretest in Survey Research," found that under certain circumstances telephone pretests were able to detect more errors than face-to-face pretests.

[27] Dillman, *Mail and Telephone Surveys.*

- Randomizing response choices. For multiple-category questions in which category sequence may influence response choice, the computer can be programmed to randomize the order of presentation of the categories *separately* for each respondent. Randomization provides neutralization of any response bias stemming from the way response categories are ordered.
- Checking for response consistency. The computer can be programmed to check for consistency between the response given to a question at hand and responses given to certain key questions asked earlier. As soon as any inconsistency is detected, the CRT screen can bring it to the respondent's attention and give him or her an opportunity to correct the inconsistency. In this way the accuracy of the collected data can be improved.
- Incorporating complex skip patterns. Earlier in the chapter we saw that complex skip patterns—such as "If the answer to question 5 is "yes" and the answer to question 7 is "no," skip to question 10; otherwise, go to question 9"—can confuse respondents and/or interviewers and hence should be avoided. However, computerized interviewing can handle even very complex skip patterns. The computer can easily check as many *if* statements and previous responses as necessary, decide which question should be asked next, and pose that question on the CRT screen almost instantaneously.
- Personalizing. Once a respondent's name is entered into the computer keyboard at the start of an interview, that name can be automatically inserted into key questions and instructions throughout the questionnaire to provide a degree of personalization that would be very time-consuming to achieve in noncomputerized interviewing. Personalization can add greatly to the rapport established with the respondent.
- Adding "new" response categories. Consider a multiple-category question with an "other" category. In computerized interviewing, when a prespecified number of respondents provide the same open-ended response to the "other" category, that response can be automatically converted into an explicit checkoff category by adding it to the set of prespecified categories. The addition of such new categories may help reduce the time required for subsequent interviews.

The preceding features are by no means an exhaustive list of the increased capabilities offered by computerized interviewing. The questionnaire designer's and computer programmer's ingenuity is the only limit on the creative ways in which computerized interviewing can be used to overcome some of the limitations of traditional interviewing. One must remember, however, that the computer cannot *design* a questionnaire; design is still a human responsibility. Computerized interviewing is therefore not a panacea for questionnaire design.

DESIGNING COVER LETTERS FOR MAIL QUESTIONNAIRES

Cover Letter *A letter to accompany a mail questionnaire in order to win the cooperation of respondents.*

Among the factors likely to have a major impact on whether a mail questionnaire is filled out and returned is the cover letter accompanying it. The primary purpose of a ***cover letter*** is to win the cooperation of respondents. Studies have shown that what the cover letter says and how it says it can affect the response rate to a mail survey.[28] Therefore a researcher planning to use a mail questionnaire must spend some time and effort on designing an effective cover letter.

A good cover letter should tell potential respondents what the study is about and, more critically, convince them of the importance of participating in it. The cover letter should also be concise and objective (i.e., it should not bias the respondent in any way). Several of the guidelines mentioned in connection with wording, appearance, and layout of questionnaires are relevant to designing cover letters as well.[29]

Exhibit 10.5 presents an illustrative cover letter and a list of its key features. The numbers identifying different sections of the cover letter correspond to the numbers accompanying the features listed. Notice that virtually all the features either directly or indirectly attempt to impress upon respondents the need to cooperate. However, not *all* these features need to be incorporated in every cover letter. In fact, a cover letter containing every listed feature can at times be self-defeating; it may become, or at least appear to be, so lengthy that some respondents may not even be interested in reading it. The purpose of Exhibit 10.5 is merely to show the variety of ways respondents can be encouraged to participate in a survey. An effective cover letter must possess at least several, but not necessarily all, of the illustrated attributes.

OPENERS FOR PERSONAL AND TELEPHONE INTERVIEWS

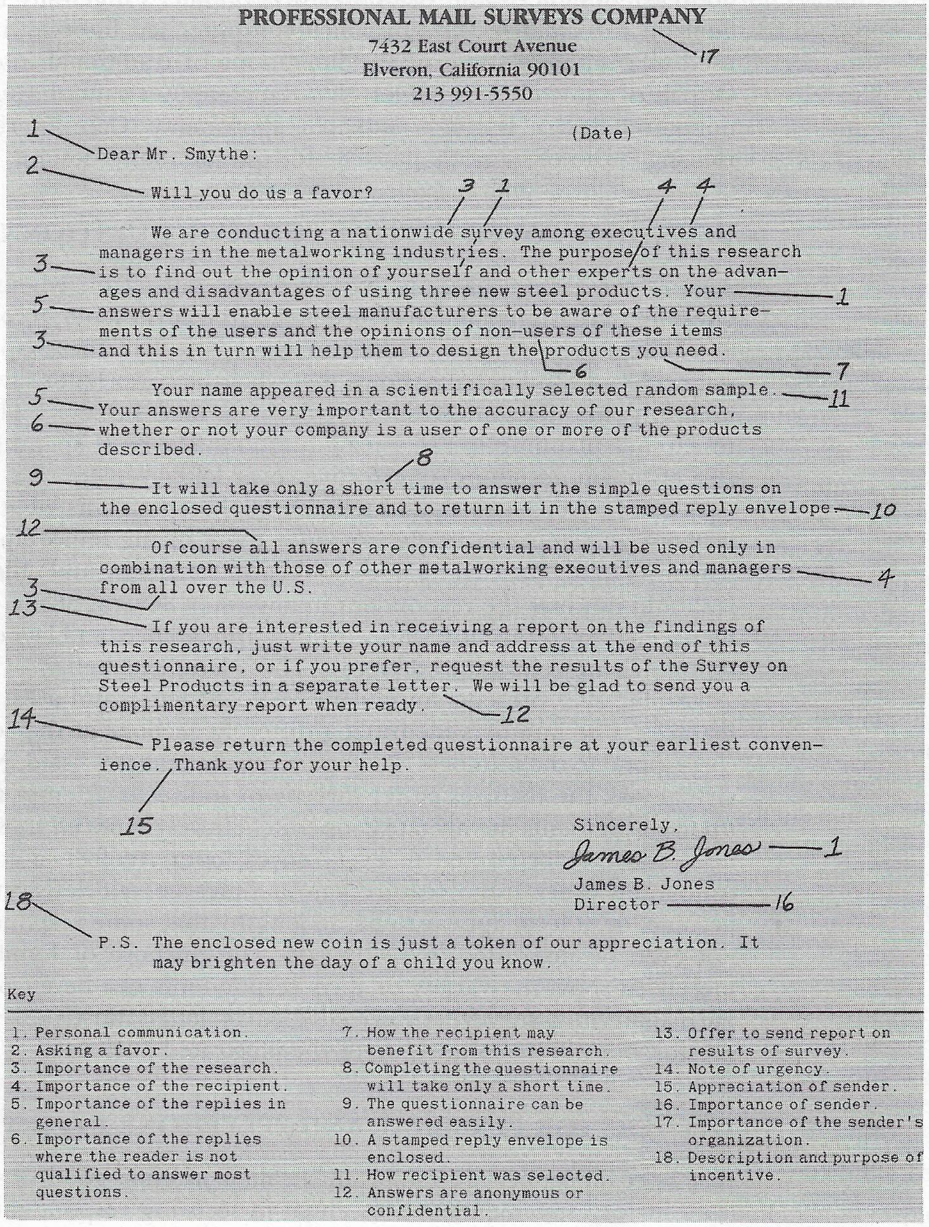
Just as cover letters are important in mail surveys, good openers (i.e., introductory statements) are essential in securing respondent coopera-

---

[28] See, for example, Michael J. Houston and John R. Nevin, "The Effects of Source and Appeal on Mail Survey Response Patterns," *Journal of Marketing Research*, 14 (August 1977), pp. 374–378.

[29] For a detailed treatment of cover letter design, see Erdos, *Professional Mail Surveys*, Chap. 12.

Exhibit 10.5 *Illustrative Cover Letter and Key Features*



PROFESSIONAL MAIL SURVEYS COMPANY

7432 East Court Avenue
Elveron, California 90101
213 991-5550

(Date)

Dear Mr. Smythe:

Will you do us a favor?

We are conducting a nationwide survey among executives and managers in the metalworking industries. The purpose of this research is to find out the opinion of yourself and other experts on the advantages and disadvantages of using three new steel products. Your answers will enable steel manufacturers to be aware of the requirements of the users and the opinions of non-users of these items and this in turn will help them to design the products you need.

Your name appeared in a scientifically selected random sample. Your answers are very important to the accuracy of our research, whether or not your company is a user of one or more of the products described.

It will take only a short time to answer the simple questions on the enclosed questionnaire and to return it in the stamped reply envelope.

Of course all answers are confidential and will be used only in combination with those of other metalworking executives and managers from all over the U.S.

If you are interested in receiving a report on the findings of this research, just write your name and address at the end of this questionnaire, or if you prefer, request the results of the Survey on Steel Products in a separate letter. We will be glad to send you a complimentary report when ready.

Please return the completed questionnaire at your earliest convenience. Thank you for your help.

Sincerely,

James B. Jones

James B. Jones
Director

P.S. The enclosed new coin is just a token of our appreciation. It may brighten the day of a child you know.

Key

1. Personal communication.
2. Asking a favor.
3. Importance of the research.
4. Importance of the recipient.
5. Importance of the replies in general.
6. Importance of the replies where the reader is not qualified to answer most questions.
7. How the recipient may benefit from this research.
8. Completing the questionnaire will take only a short time.
9. The questionnaire can be answered easily.
10. A stamped reply envelope is enclosed.
11. How recipient was selected.
12. Answers are anonymous or confidential.
13. Offer to send report on results of survey.
14. Note of urgency.
15. Appreciation of sender.
16. Importance of sender.
17. Importance of the sender's organization.
18. Description and purpose of incentive.

Source: Reproduced from Paul L. Erdos, *Professional Mail Surveys,* rev. ed. (Melbourne, Fla.: Krieger Publishing, 1983), pp. 102–103. Used by permission.

tion in personal and telephone interviews. Openers in these interviews need not, and perhaps should not, be as lengthy as a typical cover letter. But they should usually include the following features:

- An appropriate salutation, such as "Good morning! I am Ralph Johnston with Marketing Research Associates Company."
- A brief statement about the project and its purpose, such as "We are conducting a survey of heads of household regarding their feelings about modern home appliances."
- An indication of how long the interview might last, such as "This survey should take no more than ten minutes."
- A polite request for permission to conduct the interview, such as "May I please talk to the male or female head of your household for a few minutes?"

Openers for personal and telephone interviews should be carefully constructed and be an integral part of the questionnaire to ensure consistency across interviews and across interviewers. Lack of consistency in openers can result in unnecessary distortions and biases in the data to be collected. And just as a suitable opener is important for getting the interview off to a good start, common courtesy demands that the questionnaire end with an appropriate "thank you" statement.

DESIGNING OBSERVATION FORMS

As mentioned at the beginning of the chapter, designing observation forms is somewhat easier than constructing questionnaires. Nevertheless, the quality of data gathered through observation depends on the clarity of the instructions given and the tasks assigned to the observers. A standard form is not necessary when a study involves nonstructured observation; the observer simply makes mental and/or written notes about whatever is being observed. In these studies the training and the skills of the observer are critical for generating relevant and objective information.

When structured observation is involved, however, a standard form that clearly states the specific observations to be made and provides for efficient recording of those observations is desirable. Several of the guidelines concerning such aspects as wording, sequencing, and layout, which we discussed under questionnaire design, are germane to constructing observation forms as well. After all, a structured-observation form is basically a questionnaire that the observer fills out. A cluttered or confusing observation form can cause problems for the observer and result in erroneous data.

Exhibit 10.6 illustrates a form that can be used in a structured-observation study to ascertain the impact on customers of a special in-store display. The primary purpose of Exhibit 10.6 is to show what a struc-

Exhibit 10.6 *Observation Form for Recording Characteristics and Behavior of Customers Stopping at a Special Display*

Observation number ________

AS SOON AS YOU OBSERVE ANY ADULT CUSTOMER (ANY CUSTOMER WHO APPEARS TO BE 18 YEARS OF AGE OR MORE) STOPPING AT THE DISPLAY, START YOUR STOPWATCH, AND RECORD THE FOLLOWING:

1. Sex of the customer: Male ________ Female ________

2. Approximate age of the customer: 18–30 ________

31–50 ________

Over 50 ________

3. Number of individuals accompanying the customer: ________
(IF NONE, GO ON TO ITEM 4.)

a. How many of the accompanying individuals are adults? ________

b. How many are children? ________

4. Does the customer touch or handle the product?

Yes ________ No ________

5. Do any of the accompanying *adults* touch or handle the product?

Yes ________ No ________

No accompanying adult ________

6. Do the customer or accompanying adults, if any, leave the display with one or more units of the displayed product?

Yes ________ No ________
↓
How many *total* units? ________

STOP YOUR STOPWATCH WHEN THE CUSTOMER *AND* ACCOMPANYING ADULTS, IF ANY, LEAVE THE DISPLAY AREA. RECORD BELOW THE TOTAL TIME SPENT AT THE DISPLAY.

________ Minutes ________ Seconds

GO TO A NEW OBSERVATION FORM AND RECORD INFORMATION FOR THE NEXT ADULT WHO STOPS AT THE DISPLAY.

---

tured-observation form might look like. It is not to be viewed as an exhaustive list of items to be observed in order to evaluate the special display's impact. As in the case of a questionnaire, the content of an observation form—that is, the various observation tasks and items included in it—should be based on well-defined objectives and specific information requirements in a given setting.

SUMMARY

Constructing questionnaires is not an easy undertaking. Even seasoned researchers at times inadvertently overlook design flaws. Devoting adequate time and effort to questionnaire design is therefore a must. The quality and the ultimate usefulness of data to be gathered through a survey are vitally dependent on how good the questionnaire is.

No standard methodology exists for designing questionnaires; however, we can regard the design as a sequential process consisting of *interrelated* steps. Once the data requirements of a project are clearly defined, they can be transformed into a corresponding set of rough questions. The rough draft must then go through one or more iterations involving a variety of checks.

A key decision to be made concerning each question is whether it should be open-ended or structured. Open-ended questions are generally not recommended for surveys of large samples. However, they can be used as warm-up questions or as transition questions to aid the smooth flow of the interview. Structured questions can be dichotomous or multiple-category. The categories in a multiple-category question must be collectively exhaustive and mutually exclusive. One should also rotate the categories, if feasible. The number of response categories to include depends on the nature of the research objectives and the data requirements, and the type of questionnaire administration method to be used.

The relevance of each question must be carefully examined, and irrelevant questions must be dropped. This guideline is especially important since there is a normal tendency to include more questions than are really needed. Also, questions that a respondent cannot answer meaningfully or will not answer truthfully, should be eliminated if they cannot be suitably rephrased.

Question wording is perhaps the most critical determinant of data accuracy. Unfortunately, the essentials for wording questions properly cannot be reduced to a neat, concise set of rules. However, even unseasoned researchers can spot wording problems by putting themselves in the respondents' shoes and critically examining each question. Particular attention must be focused on identifying problem questions such as double-barreled, leading, one-sided, and complex questions, as well as questions with implicit assumptions.

The sequencing of questions is another determinant of data accuracy. Questions must be arranged in a logical sequence. The following principles should generally be heeded, unless certain exceptional situations demand otherwise: Questions that are personal, sensitive, or difficult to answer should be placed as far into the questionnaire as possible. Questions related to the same topic should appear together. Questions within a topic should go from general to specific (i.e., follow a funnel sequence). And question skip patterns should be simple and kept to a minimum. Questionnaire appearance and layout, if ignored, can lead to confusion and coding errors. In mail surveys they can also have a significant impact on respondents' willingness to cooperate.

Pretesting involves administering a questionnaire to a limited number of respondents to uncover and correct any faults before the final draft is prepared. A researcher's colleagues and potential users of the data to be generated are valuable critics at this stage. Resources permitting, a questionnaire should be pretested on survey respondents at least twice: once through personal interviews involving in-depth probing, and a second time using an administration procedure similar to the one proposed for the study. For pretesting to be most effective, the researcher should have already identified specific items likely to be potential trouble spots in the questionnaire and should probe for information capable of shedding light on those items.

Computerized interviewing, which is becoming increasingly popular, offers a number of questionnaire design benefits that are difficult if not impossible to achieve under conventional modes of interviewing. These benefits include capabilities to randomize response choices, check for response consistency, incorporate complex skip patterns, personalize key questions, and add "new" response choices on the basis of interviews already completed.

An element unique to mail surveys, and critical for obtaining high response rates, is the cover letter. Above all else, a good cover letter should convince respondents to cooperate and should do so in a concise and objective fashion. Personal and telephone interviews require some appropriate, concise openers, as well.

Studies involving structured observation typically use standard data collection forms. Although these forms are somewhat easier to construct than questionnaires, they must be designed with care. Several of the guidelines presented for designing questionnaires are also relevant for designing observation forms.

KEY TERMS

Questionnaire
Dichotomous question
Multiple-category question
Response category sequence rotation
Double-barreled question
Leading (loaded) question
One-sided question
Acquiescence bias (yea-saying)
Split-ballot technique
Question with implicit assumptions
Filter question

Complex question
Classification data
Funnel sequence
Inverted-funnel sequence
Pretesting
Computerized interviewing
Cover letter

QUESTIONS

1. "Designing a questionnaire for a conclusive-research project is more difficult than designing one for an exploratory-research project." Discuss this statement.
2. Using Exhibit 10.1 as a guide, briefly discuss the major ways in which a questionnaire may result in inaccurate data.
3. Give two examples to illustrate why the steps involved in questionnaire design are interrelated.
4. Under what circumstances could you use nonstructured questions in a conclusive-research study?
5. Suppose you want to ascertain the extent to which students in your class are paying for their educational expenses through their own earnings. Assuming you want to ask just one question, how would you phrase it in each of the following forms: open-ended, dichotomous, and multiple-category? In what ways would the type of data obtained through each form differ?
6. Why must responses to multiple-category questions be collectively exhaustive and mutually exclusive?
7. What is the split-ballot technique? For what purpose(s) can it be used? What are its potential drawbacks?
8. What factors should be considered in deciding on the number of categories to include in a multiple-category question?
9. Do you agree or disagree with this statement? "As long as a question pertains to at least one of the research objectives, it must be included in the questionnaire." Explain your answer.
10. Make up an example of your own to illustrate each of the following: (a) a double-barreled question; (b) a leading question; (c) a question with an implicit assumption. In each case, suggest a revised question that will reduce potential bias.
11. What is a filter question? When are such questions most appropriate?
12. What is meant by a funnel sequence? Is it always beneficial in a questionnaire? Why or why not?
13. What guidelines should one keep in mind to make pretesting as beneficial as possible?
14. State and briefly describe the questionnaire design benefits offered by computerized interviewing.
15. In what ways does the design and the use of a mail questionnaire differ from those of other types of questionnaires?

CHAPTER 11

Types of Scales & Attitude Measurement

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Define and illustrate the four levels of measurement under which numbers generated through a survey can be classified.*
- *State the appropriate central-tendency measure(s) for each level of measurement.*
- *Distinguish between attributes, behavioral variables, beliefs, and attitudes.*
- *State and illustrate the five methods for inferring people's attitudes.*
- *Describe the various dimensions on which rating scales and the nature of the data they yield can vary.*
- *Derive a rank-ordering of objects on the basis of paired-comparisons data pertaining to them.*
- *Illustrate the formats of Likert, semantic-differential, and Stapel scales.*
- *Describe how data generated by Likert, semantic-differential, and Stapel scales are analyzed and interpreted.*
- *Define validity, reliability, and sensitivity of a scale.*
- *Distinguish between different types of validity and reliability.*
- *State the reasons a multiple-item attitude scale is better than a single-item scale.*

Measurement
The quantification of observations or responses by assigning numbers to them according to a given set of rules.

A variety of question formats and their advantages and disadvantages were presented in the previous chapter. This chapter's focus is on measuring or quantifying responses to different types of questions, with particular emphasis on questions dealing with people's feelings, opinions, and the like. A typical definition of ***measurement*** is "the assignment of numbers to observations [or responses] according to some set of rules."[1] We will use this definition, although the term *measurement* has also been given somewhat different interpretations.[2]

Assigning numbers to responses has two potential benefits. First, quantified responses from a large sample can be summarized more efficiently and parsimoniously than nonquantified responses. Second, quantified responses can be manipulated by using a variety of mathematical techniques. The results of the data analyses can provide rich insights that may go undiscovered by merely examining unquantified responses.[3] Despite these attractive advantages, however, assigning numbers to survey responses in a meaningful fashion is no easy task. As will be shown in the next section, depending on the nature of the questions and the responses being quantified, the numbers assigned may not possess all the mathematical properties commonly attributed to the number system. Therefore a thorough understanding of what the scaled responses represent is essential for determining the types of analyses that can be legitimately performed by using the collected data. The next section describes four types of scales representing different levels of quantified responses and the numerical properties of each.

---

[1] Gene F. Summers, ed., *Attitude Measurement* (Chicago: Rand McNally, 1970), p. 1.

[2] Alternative definitions and interpretations of the term *measurement* are presented in Warren S. Torgerson, *Theory and Methods of Scaling* (New York: Wiley, 1958), pp. 13–14.

[3] These benefits are reflected in the remarks of Summers, *Attitude Measurement*, p. 11.

MEASUREMENT LEVELS

Let us briefly review the nature of the number system. Any given set of numbers—say, 3, 6, 9, 12, 24, and 36—has certain basic ominaisuudet properties. First, the numbers follow a *rank order;* thus 9 is greater than 3, 12 is less than 24, and so on. Second, the differences, or *intervals,* between pairs of numbers can be compared. Thus the interval between 6 and 3 is the same as the interval between 9 and 6; the interval between 36 and 24 is twice the interval between 12 and 6; and so on. This property stems from the fact that the interval separating a pair of adjacent numbers in the number system is, by definition, equal to the interval between any other such pair. Third, we can divide one number by another and Tulkita interpret the resulting suhdeluvut *ratio* as being osoittava indicative of the suhteellista suuruudet relative magnitudes of the two numbers. Thus 6 is twice as large as 3, 12 is one-third as large as 36, and so on. We can compute and interpret ratios of numbers because the number system has a *unique zero point.* In other words, all elements within the number system are measured from the *same* starting point—namely, zero.

Although the three properties of the number system are obvious and invariably taken for granted, they are of special significance in the context of quantifying survey responses. That is, to interpret quantified responses as if they were numbers is not always appropriate. Quantified responses fall into one of four *measurement levels* (also known as *scales of measurement*): nimellinen nominal, järj. ordinal, luku interval, or ratio, with varying properties. We will examine each level of measurement next.

Nominal-Scaled Responses

Nominal asteikko ***Scale*** *A set of numbers in which the numbers serve* only *solely as identification labels.*

Numbers forming a ***nominal scale*** are no more than labels and are used solely to identify different categories of responses. The following examples illustrate:

What is your sex?
1 Male
2 Female

Which one of the following media influences your purchasing decisions the most?
1 Television
2 Radio
3 Newspapers
4 Magazines

The numbers accompanying the response categories in these questions have none of the three properties described earlier. They do not imply any particular rank ordering of the responses. Moreover, the intervals between, as well as ratios of, pairs of these numbers reveal nothing about the nature of the responses. That these numbers form no more than a nominal scale is easily demonstrated by the fact that *any* set of numbers can be used to represent the response categories. For example, the following numbers would serve to describe the responses just as well as the numbers previously used:

265	Male	600	Television
575	Female	755	Radio
		523	Newspapers
		129	Magazines

sallittu

The only permissible mathematical operation with nominal-scaled responses is counting the number of responses falling within each category. From counting, one can report, for instance, that 75% of the respondents checked category 600 in answering the media question, and hence 600 is the *modal* (i.e., most frequently checked) response category. Determining any measure of ***central tendency***—a number depicting the "middle" position in a given range or distribution of numbers—other than the ***mode*** is inappropriate when response categories form only a nominal scale.

kuvaa

Central Tendency *A number depicting the "middle" position in a given range or distribution of numbers.*

Ordinal-Scaled Responses

An ***ordinal scale*** is more powerful than a nominal scale in that the numbers possess the property of *rank order*. Consider the following question:

Mode *A measure of central tendency that represents the most frequently occurring response category.*

Ordinal Scale *A set of numbers in which the magnitudes of the numbers represent no more than a rank ordering.*

How long do you spend reading newspapers on a typical weekday?
1 Less than 5 minutes.
2 5 minutes to less than 15 minutes.
3 15 minutes to less than 30 minutes.
4 30 minutes or more.

The scale values of 1, 2, 3, and 4 assigned to the response categories, in addition to serving as labels, provide an indication of the extent of newspaper reading. For instance, a respondent checking category 4 spends more time reading newspapers than one checking category 3; however, the numbers do not indicate *how much more time* the former spends than the latter. A respondent who spends 15 minutes per day and another who spends 25 minutes per day will both check category 3. Since the exact reading times cannot be inferred from the scale values, the intervals between them have no meaningful interpretation. Stated differently, given the nature of the response categories, any set of four numbers can be used as scale values, as long as they go from lowest to highest. For example, a set of scale values such as 10, 15, 25, 40 is equivalent in this case to the set 1, 2, 3, 4.

päätellä

Median *A measure of central tendency that represents the category in which the 50th percentile response falls when all responses are arranged from lowest to highest (or vice versa).*

Two measures of central tendency are meaningful for ordinal-scaled responses: their mode as well as their ***median***—that is, the category in which the 50th percentile response falls when all responses are arranged from lowest to highest (or vice versa). To illustrate, consider the following distribution of responses to the question about reading newspapers:

Response Category	Percent of Respondents Checking Category
1	40%
2	25%
3	25%
4	10%

In this case the mode is category 1 and the median is category 2.

For numbers forming an ordinal scale we can determine percentile values other than the median—for example, the 25th and 75th percentile values—so as to get a feeling for how widely or narrowly dispersed the data are on either side of the median.

Interval-Scaled Responses

Interval Scale *A set of numbers in which the differences (but not ratios) between numbers can be meaningfully interpreted.*

Interval-scaled responses are more powerful than ordinal-scaled responses because an ***interval scale*** has all the properties of an ordinal scale and, in addition, the *differences* between scale values can be meaningfully interpreted. Strictly speaking, variables such as attitudes, opinions, and preferences cannot be quantified to yield exact interval scales. Nevertheless, responses to questions like the following are frequently assumed to form an interval scale.

> How likely are you to buy a new automobile within the next six months? (Please check the most appropriate category.)
>
> Will definitely not buy ______ (1)
>
> Extremely unlikely ______ (2)
>
> Unlikely ______ (3)
>
> Likely ______ (4)
>
> Extremely likely ______ (5)
>
> Will definitely buy ______ (6)

The values assigned to this set of responses run from 1 through 6 (these numbers could also be in reverse order) and, strictly speaking, form only an ordinal scale. But they are considered to be interval-scaled under the assumption that respondents will treat the differences between adjacent response categories to be equal, especially since the categories are physically separated by equal distances.

Unfortunately, verifying whether respondents' *mental* or *psychological* perceptions of the differences between adjacent response categories are equal is very difficult, if not impossible. Some authorities have argued that scales like the one just illustrated are no more than ordinal.[4] However, others have demonstrated that such scales are robust enough to be treated as interval scales for analysis purposes.[5] Such scales are more often than not treated as having interval properties in marketing research practice.

k-a

Mean *A measure of central tendency that represents the simple average of a set of numbered responses.*

Standard Deviation *A measure of dispersion of a set of numbered responses from their mean.*

k-a

[illegible] [illegible]

Numbers forming an interval scale, in addition to possessing ordinal-scale attributes, permit one to compute their ***mean*** (their simple average) and their ***standard deviation***, which is a measure of dispersion (i.e., degree of deviation of the numbers from their mean). Suppose the question about likelihood of buying a new automobile is posed to a sample of 200 respondents and the following response distribution is obtained:

Response Category	**Number (Percent) of Respondents Checking Category**
1	10 (5%)
2	10 (5%)
3	70 (35%)
4	60 (30%)
5	20 (10%)
6	30 (15%)
	200 (100%)

The mode and the median for the distribution are 3 and 4, respectively. The mean response turns out to be 3.8, computed as follows:

$$\begin{aligned} & 1 \times .05 + 2 \times .05 + 3 \times .35 + 4 \times .3 + 5 \times .1 + 6 \times .15 \\ & = .05 + .10 + 1.05 + 1.20 + .50 + .90 \\ & = 3.80 \end{aligned}$$

[illegible] [illegible]

Thus the mean likelihood of purchase for the total sample is precisely at the 3.8 mark on the 1-to-6 scale. The computation and the interpretation of the mean are legitimate because of the assumption that the unit of measurement remains constant throughout the scale.

Now suppose the same question is administered to a *different* sample of 200—say, a sample of respondents with lower income levels than in the earlier sample—and these responses are obtained:

---

[4] See, for example, James H. Myers and W. Gregory Warner, "Semantic Properties of Selected Evaluation Adjectives," *Journal of Marketing Research*, 5 (November 1968), pp. 409–412; John A. Martilla and Davis W. Carvey, "Four Subtle Sins in Marketing Research," *Journal of Marketing*, 39 (January 1975), pp. 8–15; Gary M. Mullet, "Itemized Rating Scales: Ordinal or Interval?" *European Research*, April 1983, pp. 49–52.

[5] See, for example, Sanford Labovitz, "The Assignment of Numbers to Rank Order Categories," *American Sociological Review*, 35 (1970), pp. 515–524; Mark Traylor, "Ordinal and Interval Scaling," *Journal of the Market Research Society*, 25 (4), pp. 297–303; William D. Perreault, Jr., and Forrest W. Young, "Alternating Least Squares Optimal Scaling: Analysis of Nonmetric Data in Marketing Research," *Journal of Marketing Research*, 17 (February 1980), pp. 1–13.

Response Category	Number (Percent) of Respondents Checking Category
1	120 (60%)
2	40 (20%)
3	10 (5%)
4	10 (5%)
5	10 (5%)
6	10 (5%)
	200 (100%)

In contrast to the first sample's mean of 3.8 on the 6-point scale, the mean for the second sample is only 1.9:

$$\begin{aligned} & 1 \times .60 + 2 \times .20 + 3 \times .05 + 4 \times .05 + 5 \times .05 + 6 \times .05 \\ & = .60 + .40 + .15 + .20 + .25 + .30 \\ & = 1.90 \end{aligned}$$

Although 3.8 is twice as large as 1.9, we cannot say that the first sample is twice as likely to buy a new automobile within the next six months as the second sample. Even though the unit of measurement remains constant throughout an interval scale, its starting, or zero, point is *arbitrary*.

The ratio of two values on an interval scale is also arbitrary and has no meaningful interpretation, since it depends on the scale's starting point. To illustrate, suppose we number the six response categories 0 through 5, instead of 1 through 6. We still have an interval scale, but it has a different starting point. Exhibit 11.1 shows the impact of this change on the mean responses of the two samples. From the exhibit we see that the ratio of sample mean values on scale A is 3.8/1.9 = 2.0. The ratio of sample mean values on scale B, though, is 2.8/.9 = 3.1.

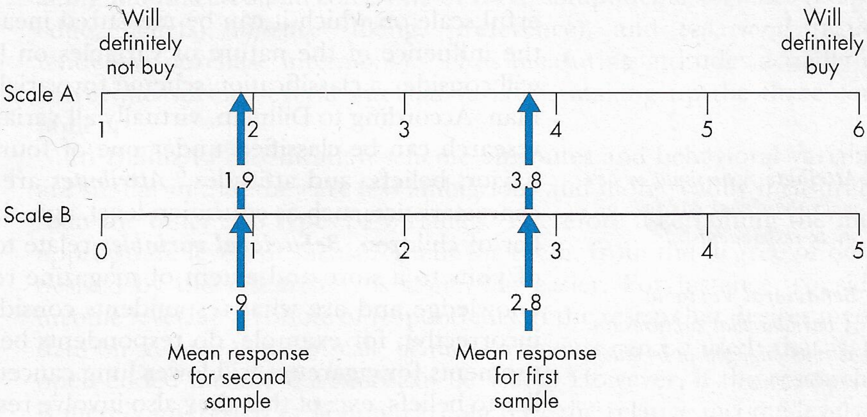
As demonstrated by Exhibit 11.1, with the new starting point for the interval scale, the first sample on the average is now over three times as likely to buy a new automobile as the second. This conclusion is obviously ludicrous, since the answers obtained from the two samples and the relative positions of their mean responses on the scale have not changed. Although this point is subtle, it is extremely important, because there is usually a strong temptation to compute and interpret ratios of quantified responses that are no more than interval-scaled. For instance, you have probably heard claims like "Users of our brand are twice as satisfied as users of competitors' brands," and "Homeowners are four times as strongly opposed to a property tax increase as renters." Such claims, even when based on data from well-conducted surveys, are suspect because scales used to measure respondents' feelings, opinions, and other internal variables are invariably only interval scales.

Ratio-Scaled Responses

Ratio Scale A set of numbers in which the ratios between numbers can be meaningfully interpreted.

Quantified responses forming a ***ratio scale*** are analytically the most versatile. They possess all the properties of the scales discussed thus far. In addition, ratios of numbers on these scales have meaningful interpre-

Exhibit 11.1 *Impact of Arbitrariness of an Interval Scale's Starting Point*



tations. Data on attributes of respondents such as attitudes and opinions typically will not satisfy the stringent requirements of a ratio scale. However, data on certain demographic or descriptive attributes, if they are obtained through open-ended questions, will have ratio-scale properties. Consider the following questions:

What is your annual income before taxes? $ ________

How far is your workplace from your home? ________ miles

Answers to these questions have a natural, unambiguous starting point, namely zero. Since the starting point is not chosen arbitrarily, as in the case of an interval scale, computing and interpreting ratios make sense. For instance, we can say that a respondent with an annual income of $40,000 earns twice as much as one with an annual income of $20,000.

In summary, quantified responses fall into one of four levels or scales of measurement: nominal, ordinal, interval, or ratio. Each scale possesses the attributes of the scales below it. That is, ratio sales have the properties of interval scales, which, in turn, have the properties of ordinal scales, and so on. The permissible data analyses and inferences depend on the scale level of the data.

One final point before we examine the classes of variables typically encountered in marketing research and the levels of measurement appropriate for them: Data with only nominal or ordinal properties are classified as ***nonmetric data***, while data with interval or ratio properties are classified as ***metric data***. The feature differentiating these two types of data is that the differences between numbers have precise meaning for metric data but not for nonmetric data.

'metrik'

Nonmetric Data *Data having only nominal- or ordinal-scale properties.*

Metric Data *Data having interval- or ratio-scale properties.*

CLASSES OF VARIABLES

The nature of a variable plays a key role in determining the most powerful scale on which it can be measured meaningfully. To demonstrate the influence of the nature of variables on levels of measurement, we will consider a classification scheme for variables proposed by Don Dillman. According to Dillman, virtually all variables encountered in survey research can be classified under one of four categories: attributes, behavior, beliefs, and attitudes.[6] ***Attributes*** are personal or demographic characteristics, such as education level, age, size of household, and number of children. ***Behavioral variables*** relate to such things as frequency of visits to a store and extent of magazine readership. ***Beliefs*** relate to knowledge and are what respondents consider to be true (correctly or incorrectly); for example, do respondents believe banning print advertisements for cigarettes will lower lung cancer deaths? ***Attitudes*** are similar to beliefs, except that they also involve respondents' evaluative judgments; for instance, do respondents feel print advertisements for cigarettes should be banned?

Attribute *A personal or demographic characteristic of study respondents.*

Behavioral Variable *A variable that captures an activity engaged in by study respondents.*

Belief *A knowledge state of study respondents reflecting what they consider to be true.*

Attitude *A person's underlying mental state capable of influencing his or her choice of actions and maintaining consistency across those actions.*

While both beliefs and attitudes pertain to a person's mental state, an attitude can be viewed as stemming from a belief (or set of beliefs) that a person might have. For instance, a person may develop a negative attitude toward cigarette advertisements on the basis of a belief that lung cancer is a major problem and a belief that banning cigarette advertisements will lower lung cancer deaths. Another person who has the same beliefs but, in addition, believes strongly that the country's constitution guarantees advertisers freedom of expression, may develop a neutral (rather than negative) attitude toward cigarette advertisements. Given that beliefs and attitudes are intertwined, formal definitions of attitudes often include beliefs as an integral component of attitudes, as discussed next.

The term *attitude* has been defined in a number of different ways.[7] The purpose here is not to present the various definitions and discuss their diversity. Rather, it is primarily to review scaling techniques suitable for measuring attitudes and, more generally, any internal disposition of people that can affect their behavior. Fortunately, even the seemingly different definitions of attitudes share certain common threads. They all depict *attitudes* as underlying mental states capable of influencing a person's choice of actions and maintaining consistency across those actions.[8] These common threads imply that attitudes can be interpreted quite broadly. Therefore the general principles underlying attitude-scaling techniques should also be useful in measuring other internal variables such as beliefs, opinions, preferences, motives, and purchase intentions. Indeed, virtually all consumer behavior textbooks portray attitudes as being multifaceted and consisting of three components: *cognitive* (knowledge, beliefs), *affective* (liking, preference), and *behavioral* (action tendency, purchase intentions).[9] Thus measuring attitudes actually involves measuring several internal variables making up the three components.

---

[6] The four classes of variables and their description are from Don A. Dillman, *Mail and Telephone Surveys: The Total Design Method* (New York: Wiley, 1978), pp. 80–86.

[7] Several of these definitions are presented in Summers, *Attitude Measurement,* pp. 1–2.

[8] For further discussion of the commonalities, see Summers, *Attitude Measurement,* pp. 1–2.

In Dillman's classification scheme attributes and behavioral variables are similar in that they are less ambiguous and more readily measurable than the other two types of variables. Therefore ascertaining the most appropriate levels of measurement for them, from the degree of detail desired by the researcher, is somewhat easier. For instance, consider income level, an attribute of respondents. If the researcher desires metric data on income, a ratio scale of measurements can be used; that is, an open-ended income question can be asked. However, if the researcher is merely interested in nonmetric data (i.e., the relative income levels of respondents), the scale of measurement can be ordinal and consist of a multiple-category income question. Similarly, consider people's smoking habits, a behavioral variable. Depending on how much detail the researcher desires, data on smoking can be obtained on a nominal scale (by asking a dichotomous question about whether or not a respondent smokes), an ordinal scale (by asking a suitable multiple-category question), or a ratio scale (by asking an open-ended question to ascertain the number of cigarettes smoked per day).

Measurement of beliefs and attitudes, however, is not quite as simple. Since beliefs and attitudes are more cognitive than factual, they are more nebulous than attributes or behavior. Therefore defining them, let alone measuring them precisely, is difficult. To understand this difficulty, let us look briefly at two types of definitions of constructs, or variables, that cannot be directly observed: constitutive, or conceptual, definitions and operational definitions.

A ***constitutive definition*** of a construct describes it in terms of other constructs. For instance, a constitutive definition of *attitude* may be "a predisposition to respond favorably or unfavorably to a stimulus object." This definition makes use of the constructs "predisposition to respond" and "stimulus object" to define *attitude*. An ***operational definition*** of a construct describes how the construct is to be measured. It lays out the steps or procedures to be completed in order to assign a value to the construct. For instance, an operational definition of a person's attitude toward a particular retail store may be as follows: It is the total of the person's expressed degree of agreement—on a 5-point, "strongly agree" to "strongly disagree" scale—with each of a set of 20 evaluative statements about various aspects of the retail store.

Constitutive (Conceptual) Definition *A definition that describes a construct in terms of other constructs.*

Operational Definition *A definition that describes a construct in terms of how it is to be measured.*

---

[9] See, for example, Carl E. Block and Kenneth J. Roering, *Essentials of Consumer Behavior* (Hinsdale, Ill.: Dryden Press, 1976), pp. 220–222.

The operational definition of a construct is guided by its constitutive definition. In the illustrative operational definition the terms "set of 20 evaluative statements" and "expressed degree of agreement" with each of those statements are, in effect, operationalizations of the constructs "stimulus object" and "predisposition to respond" in the constitutive definition. Intuitively, the accuracy of the measured attitude will depend on the degree of correspondence between the constitutive and the operational definitions—that is, the extent to which constructs in the constitutive definition are *truly* captured by the steps in the operational definition. Obviously, the more ambiguous the constructs in the constitutive definition of a variable, the more difficult it is to measure the variable precisely. For this reason obtaining accurate, unambiguous measures of unobservable variables such as beliefs and attitudes is generally difficult. Interestingly, from a measurement standpoint these variables are typically neither nominal nor ratio. They usually have ordinal or interval properties and fall in the gray area between nonmetric and metric measurements.

Consider consumers' consumption of frozen TV dinners, a behavioral variable. This variable can be measured on a nominal scale at one extreme (e.g., Have you eaten a frozen TV dinner anytime during the past month?) or on a ratio scale at the other extreme (e.g., How many frozen TV dinners have you eaten during the past month?). In terms of measuring consumers' *attitudes* toward frozen TV dinners, however, a nominal scale does not make much sense. Of course, asking a consumer "Do you have an attitude toward frozen TV dinners?" will yield a yes or no, nominal-scale response. But this question does not really measure the consumer's attitude. What it does measure is an attribute rather than an attitude. Moreover, given that everyone is likely to have *some* attitude (even if it is neutral) toward frozen TV dinners, the question is meaningless. Similarly, since there is no unambiguous, fixed zero level of attitude toward frozen TV dinners, a ratio scale of measurement is also out of the question. Therefore any meaningful measure of attitudes must be either ordinal or interval.

The complexity of attitude measurement and the resultant need for its separate treatment are succinctly described by Dillman:

> People's attitudes on a specific subject are often complex and intertwined with other attitudes and beliefs they hold. Further, attitudes are held in varying degrees and are subject to considerable fluctuation. Thus when people are asked to express their attitudes on a given subject, they are likely to engage in considerable contemplation that involves carefully considering how well the particular wording of the question reflects the degree to which they hold the attitude under investigation. . . . When a question seems likely to be sensitive to wording variations, it is advisable to ask another or perhaps several questions. Thus people interested in attitudes, and to a certain extent beliefs, often ask a number of similar questions from which they try to construct attitude scales by combining the responses

in certain ways. At the other extreme, it is seldom necessary to ask for information on attributes in more than one way.[10]

ATTITUDE SCALING

Attitudes are widely believed to be a key determinant of behavior, although some authors suggest that behavior can also lead to changes in attitudes. For instance, when a person with a mildly positive attitude toward Apple computers buys an Apple computer, his or her attitude may become much more positive following the purchase, perhaps to mentally justify the purchase. In any event, firms and institutions are constantly interested in the attitudes of various constituencies that are of interest to them. For example, General Motors may be interested in the attitudes of consumers toward its automobiles, the attitudes of employees toward its working environment, and the attitudes of shareholders toward its business performance. The United Way may be interested in the public's attitude toward charitable organizations in general and toward itself in particular. How can General Motors, the United Way, and other organizations measure the attitudes of relevant constituencies? A variety of methods are available for this purpose. Before we discuss the methods, however, we point out that attitude measurement as a whole is *indirect*. In other words, attitudes can only be inferred and cannot be directly ascertained.

Several measurable responses can offer clues about people's attitudes. Cook and Selltiz have classified them into five broad categories:

(a) Measures in which the material from which inferences johtopäät are drawn consists of self-reports of beliefs, feelings, behavior, etc., toward an object or class of objects.

(b) Measures in which inferences are drawn from observed overt avoin/julkinen behavior toward the object.

(c) Measures in which inferences are drawn from the individual's reactions to, or interpretations tulkinnat of, partially osittain structured material relevant liittyvä to the object.

(d) Measures in which inferences are drawn from performance on objective tasks where functioning may be influenced by disposition [illegible] toward the object.

(e) Measures in which inferences are drawn from physiological reactions to the object.[11]

The major focus of this chapter will be on self-report, or so-called paper-and-pencil, methods—category (a) in the list—because they are by far the most widely used in attitude-scaling studies. However, we will first briefly review the other four types of methods.[12]

---

[10] Dillman, *Mail and Telephone Surveys,* pp. 85–86, excerpted with permission from John Wiley & Sons, Inc.

[11] Stuart W. Cook and Claire Selltiz, "A Multiple-Indicator Approach to Attitude Measurement," *Psychological Bulletin,* 62 (1964), p. 38.

Observing Overt Behavior

We commonly assume that a person's behavior concerning an object will be consistent with his or her attitudes toward it. This assumption has found empirical support, although there has been debate about whether attitudes are trustworthy *predictors* of behavior.[13] Nevertheless, the notion of consistency between attitudes and behavior seems almost intuitive. For instance, a consumer with an unfavorable attitude toward a certain brand of automobile is more likely than not to buy a competing brand. A television viewer with a strong negative image of a particular firm may frown visibly upon seeing a commercial touting the firm's products. The assumed link between attitude and behavior is the rationale behind inferring the former from observing the latter.

Observation of overt behavior is useful when other attitude measurement methods are inconvenient or infeasible. For instance, an observation study can be used to ascertain the attitudes of very young children toward a variety of toys. Children in groups, or individually, can be left in a room containing the stimulus set of toys, and their behavior—how long and how enthusiastically they play with each toy—can be observed. Inferences about the children's attitudes toward each toy can be made from those observations. The use of one-way mirrors to observe focus groups (as we saw in Chapter 7) is another example of this approach to gaining insight into people's attitudes.

As you might suspect, observation of behavior can only yield rough estimates of attitudes, at best. A number of factors, other than attitudes, can influence behavior. To illustrate, a housewife may buy a cheap brand of detergent, despite an unfavorable attitude toward it, because she cannot afford a more expensive brand that she really likes. A teenager who hates cigarettes may still smoke them because of peer pressure. Thus the validity of attitudes inferred solely from overt behavior may be questionable (validity of attitude measures is discussed later). Furthermore, this approach can become quite expensive, because it requires the services of highly skilled and qualified observers.

Analyzing Reactions to Partially Structured Stimuli

As its name implies, the approach of analyzing reactions to partially structured stimuli involves asking respondents to react to, or describe in some fashion, an incomplete, vague stimulus. The responses obtained are analyzed by trained professionals to reveal the respondents' attitudes. The underlying rationale here is that a person's response to an object or situation depicted by a vague stimulus will necessarily be shaped by his or her attitudes. This approach to measuring attitudes is similar to the projective techniques discussed in Chapter 6. The various projective techniques we examined—word association test, sentence completion test, TAT, cartoon test—are examples of methods employing partially structured stimuli. Therefore these techniques can be used to infer attitudes. The advantages and disadvantages of projective techniques, as well as several illustrative applications, are given in Chapter 6 and hence will not be repeated here.

---

[12] More detailed discussions of these methods are given in Cook and Selltiz, "Multiple-Indicator Approach."

[13] See, for example, William I. Wilkie and Edgar A. Pessemier, "Issues in Marketing's Use of Multiattribute Attitude Models," *Journal of Marketing Research*, 10 (November 1973), pp. 428–441.

Evaluating Performance on Objective Tasks

In the approach of evaluating performance on objective tasks, respondents are asked to complete an ostensibly objective, well-defined task. The nature of their performance is then analyzed to infer their attitudes. To illustrate, consider the attitudes of a rural community toward attracting manufacturing firms to locate within the community. One way of using the peformance-on-objective-tasks approach to gauge the community's attitudes toward industrialization is as follows: A representative sample of residents is asked to read carefully (or even memorize) an essay containing numerous facts and figures and presenting a balanced view of the pros and cons of industrializing rural communities. After a period of time the respondents are quizzed on the essay's content. The specific parts of the essay they remember and how well they remember them will offer clues about the community's attitudes concerning industrialization. For instance, sound memory of a disproportionately large amount of unfavorable information about industrialization will signal a negative attitude toward increased industrialization.

Methods we discussed under structured, disguised questionnaires in Chapter 6 are also illustrations of the performance-on-objective-tasks approach. As implied in Chapter 6, a major drawback of this approach is the difficulty in constructing appropriate objective tasks and meaningfully interpreting performance on those tasks.

Monitoring Physiological Responses

The approach of monitoring physiological responses is based on the premise that a person's emotional reactions to a stimulus will be accompanied by corresponding involuntary physiological changes. For example, a feeling of fear or anxiety can induce physiological reactions such as shivering, perspiration, and increased heartbeat. Mechanical and electronic gadgets like the galvanic skin response (GSR) meter (to measure changes in the electrical resistance of the skin caused by changes in the amount of perspiration) and pupillometer (to measure the extent of pupil dilation in response to a visual stimulus) are available to measure physiological reactions from which internal emotions can be inferred.

Other illustrations involving the use of physiological measurements include eye-tracking equipment and response latency measures (e.g., VOPAN) that we discussed under mechanical-observation techniques in Chapter 6.

Physiological measurements, however, have a serious drawback, because they merely measure *emotional arousal* rather than *attitudes.* They cannot, by themselves, reveal whether the source of the arousal is a positive or a negative emotion. For instance, a respondent who is annoyed by a test advertisement and one who is favorably impressed by it may both be equally *interested* in it and hence display identical pupil dilations as measured by a pupillometer. Consequently, physiological responses may serve as no more than measures of the attention-getting power or excitement-generating potential of marketing stimuli such as product advertisements and packages.

One marketing research firm specializing in physiological measurements based on eye-tracking technology is Perception Research Services, Inc. The following excerpt from a brochure describing the firm's services emphasizes the inappropriateness of inferring attitudes solely on the basis of physiological measures:

> Eye tracking is an addition to, not a substitute for, the data provided by traditional question and answer procedures. Eye tracking provides an indication of two stages in the communications process:
>
> - Attention
> - Involvement
>
> To examine the remaining dimensions, it is necessary to uncover opinions, attitudes, and inclinations which reflect if the material generates a favorable impression of the brand and encourages trial and/or usage.[14]

SELF-REPORT MEASUREMENTS OF ATTITUDES

As mentioned earlier, attitudes can only be inferred and cannot be directly observed or measured. Nevertheless, self-report measures are somewhat more straightforward than the other four categories of methods. They involve asking respondents relatively direct questions concerning attitudes toward whatever is of interest to the researcher. The questions are typically in the form of rating scales on which respondents check off appropriate positions that best reflect their feelings.

Rating scales can take on a variety of physical forms. There are several key dimensions on which rating scales and the nature of the data they yield can vary: graphic versus itemized formats; comparative versus noncomparative assessments; forced versus nonforced response choices; balanced versus unbalanced response choices; labeled versus unlabeled response choices; number of scale positions; and measurement level of data obtained. We will now briefly discuss each of these dimensions, using as an illustration a scenario wherein the management of Safeway stores in a town wants to measure overall consumer attitudes toward Safeway.

---

[14] "Multidimensional Communications Research," a brochure prepared by Perception Research Services, Inc., 560 Sylvan Avenue, Englewood Cliffs, N.J. 07632, p. 1.

Graphic Versus Itemized Formats

Graphic Rating Scale
A continuous scale, presented in the form of a straight line, along which a theoretically infinite number of response choices are possible.

A ***graphic rating scale*** presents a continuum, in the form of a straight line, along which a theoretically infinite number of ratings are possible. The implicit rationale for using a graphic rating scale is that it enables one to detect fine shades of differences in attitudes. A true, or pure, graphic rating scale resembles the following question:

> Indicate your overall opinion about Safeway by placing a √ mark at an appropriate position on the line below.
>
> Very bad |————————————————————————| Very good

To quantify responses to this question, one measures the physical distance between the left extreme position and the response position on the line; the greater the distance, the more favorable is the attitude toward Safeway.

A potential difficulty in using a graphic rating scale is that coding and analysis will require a substantial amount of time since one has to first measure physical distances on the scale for each respondent. An even more serious drawback is that respondents may be incapable of even mentally perceiving fine shades of differences in attitudes, let alone accurately translating their perceptions into measurable physical distances. In other words, while graphic rating scales are capable of facilitating precise attitude measurements *in theory,* whether they can be meaningfully used *in practice* is arguable. Consequently, graphic rating scales (at least the pure variety) are not widely used in marketing research surveys.

oaifelly

Itemized Rating Scale
A scale that has a finite set of distinct response choices.

Itemized rating scales have a set of distinct response categories; any suggestion of an attitude continuum underlying the categories is implicit. They essentially take the form of the multiple-category questions discussed in Chapter 10. Typical itemized rating scale formats are illustrated by the following questions:

> Indicate your overall opinion about Safeway by checking one of the following categories.

Very bad		Bad		Neither bad nor good		Good		Very good
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

Which of the following best describes your overall opinion of Safeway?

Terrible	Poor	Fair	Good	Very good	Excellent
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

While less refined than a graphic rating scale, an itemized rating scale should be easier to respond to—and more meaningful from the respondent's perspective. Also, coding and analysis of the raw data should be somewhat less laborious. As you may expect, itemized rating scales are much more widely used than the graphic type.

To gain the benefits of both types of scales a combination format can also be used. The scale in the following question illustrates such a format.

Indicate your overall opinion about Safeway by placing a √ mark in the category that best summarizes your feelings.

Very bad ... Very good

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

The scale in this question resembles a graphic scale because its horizontal line signifies an underlying continuum. Yet the scale has a distinct set of response categories.[15]

Comparative Versus Noncomparative Assessments

Contrast the set of questions in the previous section with this set of two questions.

What is your overall rating of Safeway in comparison with other supermarkets in your area?

Much worse	Worse	About the same	Better	Much better
[ ]	[ ]	[ ]	[ ]	[ ]

Rank the following by placing a 1 beside the store you think is best overall, a 2 beside the store you think is second best, and so on.

Kroger ____
Piggly Wiggly ____
Safeway ____
Tom Thumb ____
Winn-Dixie ____

---

[15] For a discussion of other variations and properties of graphic rating scales, see, for example, Chem L. Narayana, "Graphic Positioning Scale: An Economical Instrument for Surveys," *Journal of Marketing Research,* 14 (February 1977), pp. 118–122; see also Linda W. Friedman and Hershey H. Friedman, "Comparison of Itemized Versus Graphic Rating Scales: A Validity Approach," *Journal of the Market Research Society,* 28 (July 1986), pp. 285–289.

A key difference between the two sets is that questions in the latter set *explicitly* ask respondents to compare Safeway with other stores. While the questions in the first set seek absolute or noncomparative ratings, those in the second set seek relative or comparative ratings. Which type of rating scale is better? This question has no clear-cut answer other than "It depends."

Comparative Rating Scale *A scale that seeks respondents' feelings about an issue/object by explicitly asking them to compare the issue/object with a specified frame of reference.*

Noncomparative Rating Scale *A scale that seeks respondents' feelings about an issue/object without providing any frame of reference with which to compare the issue/object.*

A ***comparative rating scale*** provides all respondents with a common frame of reference. In contrast, a ***noncomparative rating scale*** implicitly permits respondents to use any, or even no, frame of reference. A comparative rating scale, by virtue of its common frame of reference, allows the researcher to be confident that all respondents are answering the same question (recall our discussion of questions with implicit assumptions in Chapter 10). However, in the context of measuring attitudes, a standard frame of reference imposed by the researcher may not necessarily be meaningful to all respondents (e.g., some respondents may have had no exposure to or experience with the stated comparison norm). Under such circumstances the validity of the comparative ratings may be questionable.

The choice between the comparative and noncomparative formats must, therefore, be situation-specific. Depending on the nature of potential respondents and their realms of experience with the attitude objects about which ratings are desired, the researcher must decide which format is likely to be most appropriate.

Forced Versus Nonforced Response Choices

Forced-Choice Scale *A scale that does not give respondents the option of expressing a neutral opinion.*

Nonforced-Choice Scale *A scale that gives respondents the option of expressing a neutral opinion.*

A ***forced-choice scale*** does not give respondents the option of expressing a neutral, or middle-ground, attitude. A ***nonforced-choice scale*** does give respondents the option of expressing a neutral attitude. The following questions have nonforced-choice scales.

Indicate your overall opinion about Safeway by checking one of the following categories.

Very bad		Bad		Neither bad nor good		Good		Very good
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

What is your overall rating of Safeway in comparison with other supermarkets in your area?

Much worse	Worse	About the same	Better	Much better
[ ]	[ ]	[ ]	[ ]	[ ]

In general, an itemized rating scale with an odd number of response categories will have a neutral position, represented by the category falling exactly in the middle (one possible exception is a scale with unbalanced response choices, which is discussed later). By the same token, a scale with an even number of categories will typically force respondents to take a definite position on either the positive or the negative side of the scale, as illustrated by the following questions.

> Indicate your overall opinion about Safeway by placing a √ mark in the category that best summarizes your feelings.
>
> Very bad — Very good
>
> 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
>
> What is your overall rating of Safeway in comparison with other supermarkets in your area?
>
> Much worse [ ] Worse [ ] Better [ ] Much better [ ]

The issue of whether scales with forced response choices are better than those with nonforced response choices (or vice versa) has no unambiguous answer. Intuitively, a prudent choice is to include a neutral position (i.e., have an odd number of response choices) to accommodate respondents with no definite attitudes toward the subject matter. In fact, some researchers argue that even provision of an odd number of categories, with a neutral category in the middle, may not guarantee that a scale does not force respondents; they further suggest that a truly nonforced-choice scale should have a separate "not applicable" or "no opinion" category.[16] In the author's experience scales with an odd number of categories appear to be more widely used than those with an even number of categories. But provision of a neutral position may tempt certain respondents, especially those who for some reason are reluctant to reveal their true positions, to select it as a sort of fake response. Therefore scales with an even number of categories deserve serious consideration when the topic to be studied is such that few respondents are likely to have a strictly neutral attitude toward it. Just as in comparative versus noncomparative rating scales, the choice between a forced or nonforced format must be made after carefully considering the characteristics unique to the situation.

Balanced Versus Unbalanced Response Choices

The balanced-versus-unbalanced dimension is relevant in constructing itemized rating scales and relates to our discussion of one-sided questions in Chapter 10. A ***balanced scale*** is one that has an equal number of positive/favorable and negative/unfavorable response choices, as in the following questions.

Balanced Scale *A scale that has an equal number of positive/favorable and negative/unfavorable response choices.*

---

[16] See, for example, G. David Hughes, "Some Confounding Effects of Forced-Choice Scales," *Journal of Marketing Research,* 6 (May 1969), pp. 223–226.

Indicate your overall opinion about Safeway by checking one of the following categories:

Very bad		Bad		Neither bad nor good		Good		Very good
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

What is your overall rating of Safeway in comparison with other supermarkets in your area?

Much worse	Worse	About the same	Better	Much better
[ ]	[ ]	[ ]	[ ]	[ ]

Consistent with what was recommended in Chapter 10, itemized rating scales, in general, should be balanced in order to reduce response biases.

One exception to this recommendation is a situation when the true attitudes of respondents are likely to be predominantly one-sided, either positive or negative. When this situation is anticipated, one should use an ***unbalanced scale*** with a larger number of response choices on the side of the scale where the overall attitude of the respondent sample is likely to fall. For instance, if a rating scale is used to determine the importance of various automobile attributes in influencing the final purchase, respondents may not use the choices at the low-importance end of the scale as freely as those at the high-importance end. Using a balanced scale under these circumstances would lead to a phenomenon called ***end piling***, wherein most responses fall into just a few categories at one end of the scale. A scale on which end piling occurs will not be very sensitive (sensitivity of scales is described later). The following question illustrates an unbalanced rating scale that can be used if respondents' opinions about Safeway are anticipated to be predominantly positive:

Unbalanced Scale *A scale in which a majority of the choices favor one side of the issue being probed.*

End Piling *A phenomenon wherein most responses fall into just a few categories at one end of a measurement scale.*

Which of the following best describes your overall opinion of Safeway?

Terrible	Poor	Fair	Good	Very good	Excellent
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Labeled Versus Unlabeled Response Choices

Rating scales will usually have a pair of ***anchor labels*** that define their two extremes. However, a researcher has considerable latitude in deciding whether to include one or more intermediate labels and whether those labels should be in the form of words, numbers, or both.

Anchor Label *A label used to define an extremity of a measurement scale.*

No rules exist for determining the number and the types of labels to include in a scale. Intuitively, however, when generating a simple and appropriate label for an intermediate category is difficult, leaving the category unlabeled is better than making up an ill-fitting label for it. To illustrate, consider the following questions.

Indicate your overall opinion about Safeway by placing a √ mark in the category that best summarizes your feelings.

Very bad																			Very good
⊢	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	┼	⊣
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Indicate your overall opinion about Safeway by checking one of the following categories:

Very bad		Bad		Neither bad nor good		Good		Very good
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

To come up with meaningful verbal labels for the unlabeled categories in these two questions is likely to be very difficult, if not impossible; hence the best approach is not to have verbal labels for those categories. Moreover, when the researcher plans to treat the response categories as being interval-scaled (as is presumably the case in these questions), the use of inappropriate verbal labels can cast doubt on the assumption of interval data. According to one study that examined the impact of category labels, even labels such as "reasonably poor" and "reasonably good," or "pleasant" and "unpleasant," were not perceived by respondents to be equidistant from the middle "neutral" position.[17] Another study found that the types of labels used as well as their locations had an impact on responses.[18] Thus verbal labels should be used cautiously and sparingly when the quantified ratings are to have interval-scale properties (as is necessary for a number of statistical analysis techniques to be discussed in later chapters).

Numerical labels, while easier to generate than verbal labels, must also be used and interpreted cautiously. Placing numbers on a rating scale can be helpful in suggesting to respondents the notion of equal distances between adjacent response categories. Herein lies a hidden danger, however, when the scale categories already have verbal labels that suggest no more than an ordinal scale. Sequentially numbering the categories on such a scale can result in the researcher's erroneously interpreting as interval data what is really only ordinal data.

Another consideration in determining whether to use verbal labels, numerical labels, or both is the likely audience for the final research report. For instance, suppose the research report is to be presented to a group of Safeway managers and we are considering the following two alternative scales for inclusion in a survey of customers.

---

[17] Myers and Warner, "Semantic Properties," pp. 409–412.

[18] Albert R. Wildt and Michael B. Mazis, "Determinants of Scale Response: Label versus Position," *Journal of Marketing Research,* 15 (May 1978), pp. 261–267.

Indicate your overall opinion about Safeway by placing a √ mark in the category that best summarizes your feelings.

Very bad	Very good
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Which of the following best describes your overall opinion of Safeway?

Terrible	Poor	Fair	Good	Very good	Excellent
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Let us examine how we would summarize for managers the results from using these alternative rating scales. If we used the first rating scale, our summary might read something like this: "On a scale of 1 to 20 (ranging from 'very bad' to 'very good'), the average customer rating for Safeway is 14.8." If we used the second rating scale, our summary might read as follows: "Of the customers surveyed, 20% rated Safeway as 'excellent,' 40% rated it as 'very good,' 30% rated it as 'good,' 5% rated it as 'fair,' and 5% rated it as 'poor.' " Intuitively, the latter summary is likely to be more informative to the managers than the former. Therefore a scale having a small number of categories with verbal labels may at times be more appropriate than a seemingly more precise scale having a greater number of sequentially numbered categories.

One final point concerning category labels: Although our discussion has focused on verbal and numerical labels, other types of labels (e.g., picture labels) are also used to help respondents understand the categories and distinguish between them. Picture labels can be especially useful in surveying samples of children or illiterates, for instance. Exhibit 11.2 gives examples of scales with picture labels used by one marketing research firm to measure children's reactions to commercials.

Number of Scale Positions

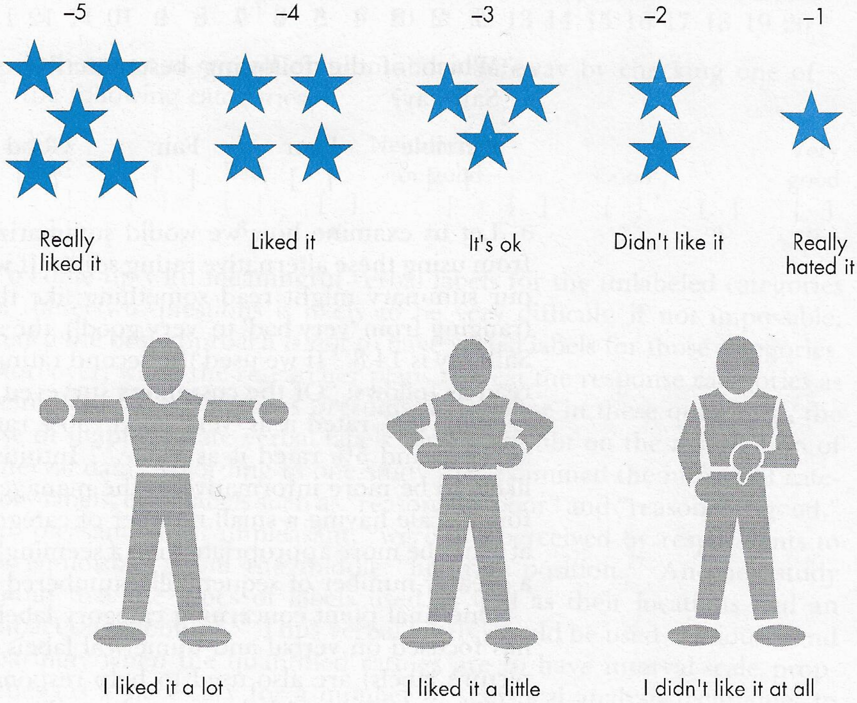
The number of points or categories to include in a rating scale is another area with no rigid rules. However, rating scales used in most surveys typically have between five and nine categories. Logically, more precise measurements should result as the number of scale positions increase.[19] Of course, a scale with a large number of positions will not be meaningful if respondents are unable to make fine mental distinctions with respect to whatever is being measured. Therefore the nature of the vari-

---

[19] This intuition is supported by empirical studies examining the impact of the number of scale positions on the nature of the information obtained; see, for example, Warren S. Martin, "Effects of Scaling on the Correlation Coefficient: Additional Considerations," *Journal of Marketing Research,* 15 (May 1978), pp. 304–308; see also Gilbert A. Churchill, Jr., and J. Paul Peter, "Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis," *Journal of Marketing Research,* 21 (November 1984), pp. 360–375.

Exhibit 11.2 *Rating Scales with Picture Labels*

How much did you like what you just saw?
(Cross out one answer.)



Source: Adapted from "The Child Research Services," a brochure prepared by McCollum/Spielman Associates, Inc., 18 East 48th Street, New York, N.Y. 10017. Used by permission.

able as well as the capabilities of potential respondents must be taken into account when one is deciding on the number of scale positions to use.

Another factor to be considered is whether the variable will be measured by a single-item scale represented by just one question (such as the ones we have seen thus far) or by summing response values obtained from a number of questions, that is, by using a multiple-item scale (we will discuss multiple-item scales in a subsequent section). The literature suggests that for accurate measurements a larger number of scale positions are needed when a single-item scale is used than when a multiple-item scale is used.[20]

---

[20] A discussion of this issue is in Jum C. Nunnally, *Psychometric Theory* (New York: McGraw-Hill, 1967), Chap. 14.

Measurement Level of Data Obtained

Measurement Level *The extent of precision of survey data in terms of whether they form a nominal, ordinal, interval, or ratio scale.*

Measurement level refers to how powerful the data are—whether they are nominal, ordinal, interval, or ratio. The type of question and the type of rating scale used have a major bearing on the measurement level of the data generated. To illustrate this point, let us consider the questions and rating scales in Exhibit 11.3. In this exhibit, questions 1 through 6 are ones we have seen while discussing the previous dimensions of rating scales. Questions 7 and 8 are new and have been included to illustrate other available options for ascertaining customers' perceptions of Safeway.

Constant-Sum Scale *A scale that asks respondents to allocate a fixed number of points to various response categories in proportion to how the respondents feel about those categories (e.g., how important each category is or how much they like each category).*

Let us first consider question 8, asking respondents to allocate 100 points among the five stores. The scale implied in this question is commonly known as a ***constant-sum scale***, and it has a natural starting point (namely, zero). In question 8 the lowest possible rating for each store is zero, which is nonarbitrary and remains the same across all respondents. Constant-sum scales, by virtue of their fixed starting point, have ratio-scale properties. Thus a respondent's opinion of Safeway can be said to be twice as favorable as his or her opinion of Kroger if he or she allocates, say, 50 points to Safeway and 25 points to Kroger. The graphic rating scale in question 1 also has ratio-scale properties, since responses on it are quantified as physical distances (e.g., inches), which necessarily have a fixed zero point.

Now examine question 7, asking respondents to select one store from each of several pairs of stores. What measurement level is achieved in this case? For each pair of stores a respondent can be classified as either one who "prefers Safeway" or one who "does not prefer Safeway." At first glance this classification resembles a nominal scale. However, the two categories are actually ordinal; the "prefers Safeway" category connotes a more favorable attitude toward Safeway than the "does not prefer Safeway" category.

Paired-Comparison Rating Scale *A scale that seeks comparative evaluations of two issues/objects at a time.*

Items seeking comparative evaluations of two objects at a time are sometimes called ***paired-comparison rating scales***. These scales are frequently used in marketing research, especially in studies involving consumer evaluations of physical products or product concepts.[21] While a respondent's task in comparing two objects at a time is seemingly simple, using paired-comparison rating scales to assess attitudes toward each of several objects has a serious drawback: Even when the number of objects to be evaluated is modest, a disproportionately large number of pairs have to be compared. For example, assessment of 6 objects requires 15

---

[21] Paired-comparison ratings are also used as input for an analysis technique called multidimensional scaling. This technique is capable of inferring metric distances between objects from an analysis of ordinal ratings. It is discussed briefly in Chapter 19. Further details of it are available in Paul E. Green, "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing*, 39 (January 1975), pp. 24–31; Paul E. Green and M. G. Greenberg, "Ordinal Methods in Multidimensional Scaling and Data Analysis," in *Handbook of Marketing Research*, ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 3.44–3.61.

Exhibit 11.3 *Different Types of Rating Scales*

1. Indicate your overall opinion about Safeway by placing a √ mark at an appropriate position on the line below.

Very bad		Very good
|———————————————————————|		

2. Indicate your overall opinion about Safeway by placing a √ mark in the category that best summarizes your feelings.

Very bad																			Very good
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

3. Indicate your overall opinion about Safeway by checking one of the following categories:

Very bad		Bad		Neither bad nor good		Good		Very good
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
1	2	3	4	5	6	7	8	9

4. Which of the following best describes your overall opinion of Safeway?

Terrible	Poor	Fair	Good	Very good	Excellent
[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

5. What is your overall rating of Safeway in comparison with other supermarkets in your area?

Much worse	Worse	About the same	Better	Much better
[ ]	[ ]	[ ]	[ ]	[ ]

6. Rank the following by placing a 1 beside the store you think is best overall, a 2 beside the store you think is second best, and so on.

Kroger ____
Piggly Wiggly ____
Safeway ____
Tom Thumb ____
Winn-Dixie ____

7. In each of the following pairs, which store do you think is better? (Please check one store within each pair.)

____ Kroger or ____ Safeway
____ Safeway or ____ Piggly Wiggly
____ Tom Thumb or ____ Safeway
____ Safeway or ____ Winn-Dixie

8. Allocate a total of 100 points among the following stores, depending on how favorable you feel toward each; the more highly you think of each store, the

(Continued)

Exhibit 11.3 *Continued*

more points you should allocate to it. (Please check that the allocated points add to 100.)

Kroger ____ points
Piggly Wiggly ____ points
Safeway ____ points
Tom Thumb ____ points
Winn-Dixie ____ points
100

paired comparisons. And the number of paired comparisons rises sharply even with a slight increase in the number of objects. [When the number of objects to be evaluated is n, the number of possible paired comparisons is given by $n(n - 1)/2$.] However, for a small number of objects the paired-comparison approach can be easily implemented to derive a rank ordering of the objects. Research in Use 11.1 illustrates how this can be done.

Let us now examine the measurement levels implied by the remaining questions (2 through 6) in Exhibit 11.3. Data to be generated by question 6 will clearly be ordinal since the numerical responses in this case, by definition, are no more than ranks. The measurement levels implied by questions 2 through 5, strictly speaking, fall in the gray area between ordinal and interval scales. However, in practice they are almost always assumed to have interval properties.[22]

In summary, the foregoing discussion, while describing the number of ways rating scale formats can vary, does not recommend any particular format as being the most desirable. Indeed, there is no such thing as *the* ideal scale format. On the basis of a controlled study examining the appropriateness of several different rating scales, one researcher concluded: "No one scale is best for all products or all types of studies."[23] Thus the choice of a rating scale should be made after one carefully considers the characteristics and requirements unique to the research setting—the nature of the variable to be measured, the extent to which potential respondents are capable of making refined mental judgments concerning the variable, and the types of analyses the researcher intends to perform on the data to be collected.

---

[22] Such an assumption is supported by the results of several studies that have found no material difference in the conclusions when ordinal data are treated as such and when they are assumed to have interval properties. See, for example, Sanford Labovitz, "The Assignment of Numbers to Rank Order Categories," *American Sociological Review,* 35 (1970), pp. 515–524.

[23] Jack Abrams, "An Evaluation of Alternative Rating Devices for Consumer Research," *Journal of Marketing Research,* 3 (May 1966), p. 191.

Research in Use 11.1 *The Coca-Cola Company*

Deriving a Rank Order from Paired-Comparison Ratings

The Coca-Cola Company wanted to rank-order the following four brands of cola in terms of taste preferences of consumers: current formulation of Coca-Cola (labeled C), a new formulation of Coca-Cola (labeled N), Pepsi (labeled P), and RC Cola (labeled R). A representative sample of 1000 consumers took part in a blind taste test wherein the colas were only identified by their labels. A paired-comparison procedure was used in the taste test—i.e., each consumer was presented with six different pairs of colas (representing all possible combinations of the four colas), one pair at a time. After tasting the two brands in each pair, the consumer was asked to express a preference for one over the other. The results of this taste test were as follows:

	Number of Customers Preferring	
Paired Brands	**First Brand in Pair**	**Second Brand in Pair**
C versus N	400 [C]	600 [N]
N versus P	500 [N]	500 [P]
P versus R	700 [P]	300 [R]
R versus C	300 [R]	700 [C]
R versus N	200 [R]	800 [N]
P versus C	400 [P]	600 [C]

number of paired comparisons in which C was preferred =
400 + 700 + 600 = 1700
number of paired comparisons in which N was preferred =
500 + 600 + 800 = 1900
number of paired comparisons in which P was preferred =
700 + 400 + 500 = 1600
number of paired comparisons in which R was preferred =
300 + 200 + 300 = 800
total number of paired comparisons = 6000

The proportion of the total number of paired comparisons in which a brand was the preferred choice is an indicator of the relative preference or "rank" for that brand. Since the total number of paired comparisons is a constant (i.e., 6000), the preference ranks can be inferred directly from the number of paired comparisons in which each brand was the preferred brand. Therefore, on the basis of the taste test results, one can infer that the most preferred brand is N, the second most preferred brand is C, the third most preferred brand is P, and the least preferred brand is R.

Note: The numbers used in this illustration are hypothetical.

COMMONLY USED MULTIPLE-ITEM SCALES

Single-Item Scale *A scale that consists of just one question pertaining to the construct it is intended to measure.*

Multiple-Item Scale *A scale that consists of two or more questions pertaining to the construct it is intended to measure.*

A ***single-item scale***, as its name implies, attempts to measure feelings through just one rating scale. For example, the rating scales in questions 1 through 5 in Exhibit 11.3, used separately, are single-item scales. In contrast, a ***multiple-item scale*** is one that contains a number of statements pertaining to the attitude object, each with a rating scale attached to it; the *combined rating*, usually obtained by summing the ratings on the individual items, is treated as a measure of attitude toward the object. Multiple-item scales are frequently used in marketing research to measure attitudes.

Developing a multiple-item scale capable of providing sound measures of attitudes toward something is no simple task. Pioneering work in establishing a rigorous procedure for constructing multiple-item attitude scales was done by L. L. Thurstone, and the general format of the resulting scale has been named after him—the Thurstone equal-appearing intervals scale, or simply the Thurstone scale, as it is often called. The procedure involved in developing a Thurstone scale is quite complex and will not be discussed here.[24] Furthermore, some questions have been raised about interpreting the data generated through a Thurstone scale, and respondents have been known to experience difficulty in meaningfully expressing their views on such a scale.[25] Perhaps as a result of these limitations, the Thurstone scale is only very rarely used these days.[26] Nevertheless, we must point out that several of the ideas involved in constructing and using the more popular attitude scales—three of which will be discussed in this section—have their roots in the original Thurstone-scaling procedure.

Likert Scale

Likert Scale *A multiple-item scale consisting of a series of evaluative statements concerning an issue or object; responses to each statement are provided on a five-point, agree-disagree scale.*

Based on a format originally developed by Rensis Likert,[27] the ***Likert scale*** consists of a series of evaluative statements (or items) concerning an attitude object. Each statement has a five-point agree-disagree scale. The number of statements included in the scale may vary from study to study, depending on how many relevant characteristics the attitude object has. However, a typical Likert scale has about twenty to thirty statements.

---

[24] The method is discussed in detail in L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (Chicago: University of Chicago Press, 1929).

[25] G. David Hughes, "Selecting Scales to Measure Attitude Change," *Journal of Marketing Research,* (February 1967), pp. 85–87.

[26] According to one survey of marketing research departments in business firms, only about 7% reported using Thurstone scales. See Barnett A. Greenberg, Jac L. Goldstucker, and Danny N. Bellenger, "What Techniques Are Used by Marketing Researchers in Business?" *Journal of Marketing*, 41 (April 1977), pp. 62–68.

[27] Rensis Likert, "A Technique for the Measurement of Attitudes," in *Attitude Measurement*, ed. Gene F. Summers (Chicago: Rand McNally, 1970), pp. 149–158.

Exhibit 11.4 *Likert Scale Items*

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1. Clerks at the store's + checkouts are friendly	____	____	____	____	____
2. Lines at the store's − checkouts move slowly	____	____	____	____	____
3. Prices in the store are + reasonable	____	____	____	____	____
4. The store has a wide variety of products to + choose from	____	____	____	____	____
5. The store's operating − hours are inconvenient	____	____	____	____	____
6. The store has a − rather confusing layout	____	____	____	____	____

Exhibit 11.4 presents six illustrative statements from a Likert scale that a supermarket like Safeway can use to measure customer attitudes toward itself and its competitors. Two points concerning the statements in Exhibit 11.4 are worthy of special note. First, the response categories have verbal labels but no numerical labels; this format is the usual format in which Likert scale statements are presented to respondents. After the scale is administered, however, numbers are assigned to the responses in order to generate a quantified measure of attitudes. The numbers 1 through 5 are normally employed for this purpose, although other sets of numbers (e.g., +2, +1, 0, −1, −2) can also be used.

Second, while statements 1, 3, and 4 are favorable toward the store, the others are unfavorable—another typical feature of statements making up a Likert scale. A good Likert scale must have a balanced set of statements, containing approximately the same number of favorable and unfavorable statements. Such a mix of statements will reduce the chances of [illegible] acquiescence bias (discussed in Chapter 10), which can occur if all statements are in the same direction.

Typical instructions for responding to a Likert scale—say one that pertains to Safeway—are as follows:

Instructions. For each of the following statements, check one of the five categories depending on how strongly you feel the statement applies to Safeway. If you strongly agree that the statement is descriptive of Safeway, check the rightmost category. If you strongly disagree that the statement is descriptive of Safeway, check the leftmost category. If your feelings are less strong, check a category between the two extremes that best reflects your feelings.

A respondent's overall attitude is measured by summing his or her numerical ratings on the statements making up the scale. However, since some statements will be favorable and others unfavorable, there is one important task to be performed before summing the ratings; namely, numbers must be assigned to the scale categories in such a way that a high (or low) numerical rating on each statement always represents the *same* attitude direction. In other words, the "strongly agree" category attached to favorable statements and the "strongly disagree" category attached to unfavorable statements must both be assigned the same number (e.g., the number 5, if a 1-through-5 numbering scheme is used).

The overall usefulness of a Likert scale depends on how carefully the statements making up the final scale are chosen. The statements must be sufficiently varied to capture all relevant aspects of the attitude object. In addition, they must be unambiguous to minimize erratic fluctuations in the responses. Last, but by no means least, each statement must be sensitive enough to discriminate between respondents with differing attitudes. For instance, consider the statement: "Safeway is a very large supermarket chain in the United States." Customers with favorable attitudes toward Safeway, as well as those with unfavorable attitudes, are bound to agree strongly with this statement. Consequently, such statements must be eliminated from the final scale, since they will contribute little toward the measurement of true attitudes.

levelin/ monnaina vallata nabo barwat
wbsnoeit

As implied by the foregoing discussion, designing a good Likert scale involves first generating a large pool of statements relevant to the measurement of an attitude and then eliminating from the pool statements that are vague and/or nondiscriminating.[28] While a comprehensive discussion of this process is beyond our scope, let us look at an example to gain an intuitive understanding of it.

Suppose Safeway wants to develop a 20-item scale to measure customer attitudes toward different supermarkets. The first step is to develop a large number of items, say 100, similar to the six statements shown in Exhibit 11.4. There are no rules for generating the initial pool of items, except that they should reflect the entire range of factors likely to influence customer attitudes. Managerial judgment and exploratory research, such as informal discussions with supermarket employees and customer focus groups, can be helpful in generating a broad range of items. A procedure similar to the one outlined next can be used in reducing the initial 100-item pool to a final 20-item instrument.

---

[28] General procedures for generating an initial item pool and then narrowing it are discussed by Gilbert A. Churchill, Jr., "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research,* 16 (February 1979), pp. 64–73; see also David W. Gerbing and James C. Anderson, "An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment," *Journal of Marketing Research,* 25 (May 1988), pp. 186–192. For a specific illustration of how a Likert scale is developed, see William J. Lundstrom and Lawrence M. Lamont, "The Development of a Scale to Measure Consumer Discontent," *Journal of Marketing Research,* 8 (November 1976), pp. 373–381.

A questionnaire containing the initial 100 items, each with five possible response categories ranging from "strongly agree" to "strongly disagree," is administered to a representative sample of customers. These customers are asked to rate Safeway (or any other supermarket) on each statement by checking one of the five response categories. The categories are then numbered 1 through 5, with higher numbers reflecting more favorable perceptions. Since there are 100 items, a respondent's total score (summed across all items) can range from 100 to 500. Further, since the initial item pool was designed to reflect various aspects relevant to customer attitudes, the total score can be viewed as a rough indicator of the respondent's overall attitude—the higher the score, the more favorable is the attitude. This total score plays a crucial role in condensing the initial pool of items into an appropriate concise scale.

To illustrate, consider the following total scores of two respondents, A and B, as well as their scores on two individual items, i and j:

Respondent	**Score on item i**	**Score in item j**	**Total Score**
A	3	4	428
B	3	1	256

On the basis of the total scores, A has a more favorable attitude than B. By looking at their scores on items i and j, we see that item i is a poor discriminator and item j is a good discriminator between the two respondents. Item j is thus likely to be a better indicator of attitudes and hence more useful in a multiple-item scale. If we extend this insight across all statements and respondents, we see that items on which respondents' scores correlate strongly with the respondents' total scores are better candidates for inclusion in the final scale than other items. Therefore of the 100 initial items, the 20 items with the highest correlations between individual scores and total scores are chosen to form the final scale.

This scale development process can be quite laborious if it is to be done in an organized fashion. Nevertheless, such a systematic approach is essential for constructing a valid, reliable, and sensitive Likert scale.

The final scale can be used to measure customer attitude toward different supermarkets, as follows: A questionnaire containing the 20-item scale is administered to an appropriate sample of respondents, and their total scores are computed. Since there are 20 items, each with a five-point scale, the total scores will range from 20 to 100; the higher the respondents' scores, the more favorable are their attitudes. Also, the average total scores (across all respondents) for the various supermarkets can be compared to get a feeling for customers' relative attitudes toward

Exhibit 11.5 *Semantic-Differential Scale Items*

1. Friendly checkout clerks	___:___:___:___:___:___:___	Unfriendly checkout clerks
2. Slow-moving checkout lines	___:___:___:___:___:___:___	Fast-moving checkout lines
3. Low prices	___:___:___:___:___:___:___	High prices
4. Wide product assortment	___:___:___:___:___:___:___	Narrow product assortment
5. Inconvenient store hours	___:___:___:___:___:___:___	Convenient store hours
6. Confusing store layout	___:___:___:___:___:___:___	Clear store layout

them. For instance, if supermarkets XYZ and ABC have average total scores of 40 and 80, respectively, we can infer that customer attitude toward ABC is more favorable than that toward XYZ. However, we *cannot* infer that customer attitude toward ABC is *twice* as favorable as that toward XYZ, because the scores on each scale item, and hence the total scores, have only interval-scale properties. The 20-item scale can also be used to monitor customer attitudes toward different supermarkets over time by administering it to representative samples of customers at different times and examining the changes in average attitude scores for each supermarket.

Semantic-Differential Scale

The format for the semantic-differential scale originated from work done some three decades ago by Osgood, Suci, and Tannenbaum to investigate the perceived meanings of words and concepts.[29] The initial semantic-differential-scaling procedure has since been modified by marketing researchers to measure consumer attitudes.[30] The ***semantic-differential scale*** is similar to the Likert scale in that it consists of a series of items to be rated by respondents. However, there are some key differences, as the six semantic-differential scale items in Exhibit 11.5 demonstrate.

Semantic-Differential Scale *A multiple-item scale consisting of a series of bipolar adjectival words or phrases pertaining to an issue or object; responses to each pair of opposite adjectives are provided on a seven-category scale separating the adjectives (with no intermediate numerical or verbal labels).*

Exhibit 11.5 highlights three basic features of a typical semantic-differential scale:

1. It consists of a series of bipolar adjectival words or phrases (rather than complete statements, as in the case of a Likert scale) that pertain to the attitude object.

---

[29] Details about this work are given in Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, *The Measurement of Meaning* (Urbana, Ill.: University of Illinois Press, 1957).

[30] See, for example, William A. Mindak, "Fitting the Semantic Differential to the Marketing Problem," *Journal of Marketing*, 25 (April 1961), pp. 29–33.

2. Each pair of opposite adjectives is separated by a seven-category scale, with neither numerical labels nor verbal labels other than the anchor labels.
3. While some individual scales have favorable descriptors on the right-hand side, the other scales are reversed, with favorable descriptors appearing on the left-hand side. The rationale for the reversals is similar to the rationale for having a mixture of favorable and unfavorable statements in a Likert scale.

Typical instructions for responding to a semantic-differential scale—say one that pertains to Safeway—are as follows:

> **Instructions.** Listed below are a number of rating scales, each containing a range of seven categories separated by a pair of opposite descriptive phrases. For each scale, check one category that best summarizes your feelings about Safeway. If you feel that one of the two phrases is a totally accurate description of Safeway, check the category closest to that phrase. If you feel that neither phrase is a totally accurate description of Safeway, check a category between the two extremes that best reflects your feelings.

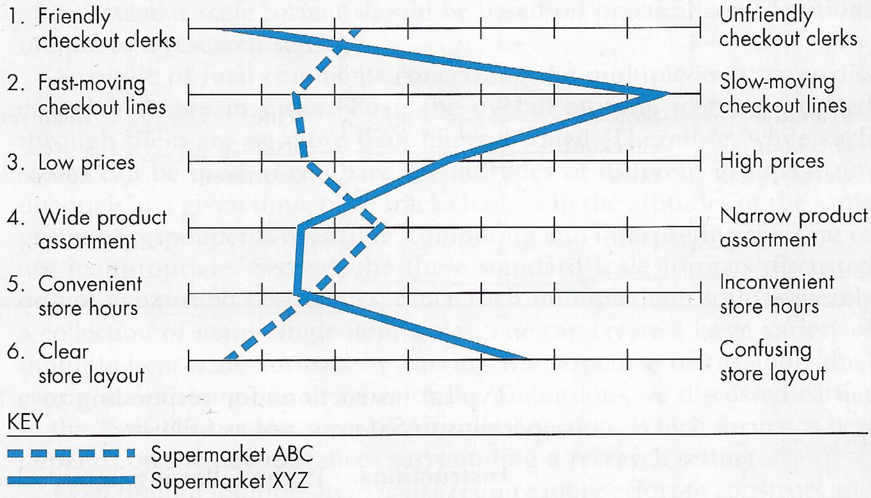
An overall attitude score for each respondent can be computed through a procedure similar to the one used with Likert scales. The seven categories can be numerically coded, say 1 through 7 (making sure to reverse code items that are reversed), and overall attitude scores can be obtained by summing the coded responses on the individual items. The interpretation of these scores is similar to the interpretation of scores generated through a Likert scale.

A more common application of the semantic-differential scale, however, is to develop a *pictorial profile* of the attitude object(s) based on the mean ratings on the individual items. Exhibit 11.6 presents hypothetical pictorial profiles for two supermarkets. To facilitate interpretation of the profiles, one usually places all the favorable descriptors on the same side of these diagrams.

According to Exhibit 11.6 customers apparently have a more favorable overall attitude toward supermarket ABC than toward supermarket XYZ. Notice that the overall profile for ABC is in general closer to the left end (i.e., more favorable end) of the scales. When we look at specific items, we see that ABC has a significant edge over XYZ with respect to speed of checkout lines and store layout; it is also perceived to be superior with respect to prices. Store XYZ is perceived to be better than ABC in terms of friendliness of checkout clerks, product assortment, and operating hours; however, ABC does not appear to be at a significant disadvantage along these dimensions.

The semantic-differential scale has considerable practical appeal because profiles like the ones in Exhibit 11.6 can pinpoint a particular

Exhibit 11.6 *Pictorial Profiles Based on Semantic-Differential Ratings*



firm's relative strengths and weaknesses as perceived by customers. The profiles can thus have useful, immediate managerial implications. Not surprisingly, the semantic differential is the most widely used attitude-scaling technique according to one recent survey.[31] Generating a useful and comprehensive set of adjectival pairs necessary for an effective semantic-differential scale is not always as easy as it may seem. The constraints and caveats mentioned in the context of developing effective Likert scales are relevant here as well.

Stapel Scale

Stapel Scale *A multiple-item scale consisting of a series of words or phrases pertaining to an issue or object; responses to each item are provided on a 10-category scale containing numerical labels for the categories.*

The ***Stapel scale*** is really a variation of the semantic-differential scale. The format of a typical stapel scale is illustrated by the six items presented in Exhibit 11.7. Notice that the items in Exhibits 11.4, 11.5, and 11.7 correspond to the same set of supermarket traits. A comparison of these exhibits should reveal the key format differences between the Likert, semantic-differential, and Stapel scales.

The Stapel scale has four distinctive features:

1. Each item has only one word or phrase indicating the dimension it represents.
2. Each item has 10 response categories.
3. Each item is a forced-choice scale since it has an even number of categories.
4. The response categories have numerical labels but no verbal labels.

---

[31] Greenberg, Goldstucker, and Bellenger, "What Techniques Are Used?"

Exhibit 11.7 *Stapel Scale Items*

Friendly checkout clerks	Slow checkout lines	Low prices	Wide product assortment	Inconvenient store hours	Confusing store layout
+5	+5	+5	+5	+5	+5
+4	+4	+4	+4	+4	+4
+3	+3	+3	+3	+3	+3
+2	+2	+2	+2	+2	+2
+1	+1	+1	+1	+1	+1
−1	−1	−1	−1	−1	−1
−2	−2	−2	−2	−2	−2
−3	−3	−3	−3	−3	−3
−4	−4	−4	−4	−4	−4
−5	−5	−5	−5	−5	−5

Typical instructions for responding to a Stapel scale, say, one that pertains to Safeway, are as follows:

> **Instructions.** In the following items, select a *plus* number for words that you think describe Safeway accurately. The more accurately you think the word describes Safeway, the larger the *plus* number you should choose. Select a *minus* number for words you think do not describe Safeway accurately. The less accurately you think a word describes Safeway, the larger the *minus* number you should choose. Therefore you can select any number from +5, for words you think are very accurate, to −5, for words you think are very inaccurate.[32]

Data obtained through Stapel scales can be analyzed by using procedures similar to the ones we discussed under semantic-differential scales. Overall attitude scores can be computed for the respondents by summing their ratings on the individual items. Alternatively, pictorial profiles of the attitude objects can also be constructed from the mean (or median) respondent ratings on each item.

One apparent advantage of the Stapel scale is that one does not have to develop complete statements or come up with pairs of bipolar words or phrases—tasks that can get to be very tedious. Nevertheless, the Stapel scale is not as widely used as the other two types, perhaps because its format and instructions appear to be relatively more complex. However, this reason for the infrequent use of the Stapel scale is speculative. In fact, one controlled study found no significant differences between the Stapel and semantic-differential formats in terms of ease of administration and the quantified attitude estimates obtained.[33] According to another study, attitude measures generated by Likert, semantic-differential, and Stapel scales were all quite similar.[34] Therefore the final choice of a particular scale format should be based on practical considerations unique to a research setting.

---

[32] Adapted from Irving Crespi, "Use of Scaling Technique in Surveys," *Journal of Marketing*, 25 (July 1961), p. 71.

[33] Del I. Hawkins, Gerald Albaum, and Roger Best, "Stapel Scale or Semantic Differential in Marketing Research?" *Journal of Marketing Research*, 11 (August 1974), pp. 318–322.

A couple of final comments concerning the multiple-item scales discussed here are in order. First, the overall attitude scores generated through them are no more than interval-scaled. Therefore, while such scores can be used to compare the attitudes of different groups of individuals at a given time, or to track changes in the attitudes of the same group of respondents over time, computing and interpreting their ratios are inappropriate. Second, the three standard scale formats discussed do not exhaust all possibilities. Since each multiple-item scale is merely a collection of many single-item scales, one can create a large variety of multiple-item scale formats by varying the structure of the individual rating scales along one or more of the dimensions we discussed earlier in the "Self-Report Measures of Attitudes" section. Which format is best depends on the circumstances surrounding a research setting.

Even though multiple-item scales require more effort to construct and use, they are capable of producing much better measures of attitudes than single-item scales. The next section briefly examines the key strengths of multiple-item scales from a measurement standpoint.

STRENGTHS OF MULTIPLE-ITEM SCALES

A single-item scale, while capable of offering clues about respondents' overall feelings toward an object, is a rather crude measure. To understand why, we must consider three criteria on which the adequacy of attitude scales should be judged: validity, reliability, and sensitivity. We will now define and discuss these criteria and indicate why multiple-item scales are better than single-item scales along each criterion.

Validity

Validity *A criterion for evaluating measurement scales; it represents the extent to which a scale is a true reflection of the underlying variable or construct it is attempting to measure.*

The ***validity*** of a scale is the extent to which it is a true reflection of the underlying variable it is attempting to measure. Alternatively, it is the extent to which the scale fully captures all aspects of the construct to be measured. Single-item attitude scales are typically deficient on this criterion because, as we have already seen, attitudes are multifaceted; a number of different factors concerning an object or issue can contribute to how a person feels about it. For instance, one's attitude toward Safeway is likely to be shaped by what one knows and feels about Safeway's appearance, physical layout, product assortment, employees, and other such factors. When a researcher attempts to ascertain respondents' attitudes solely through a single-item rating scale, there is invariably doubt about whether the scale is trustworthy. Therefore a carefully designed multiple-item scale containing numerous items, each intended to tap the respondent's position on a key facet of the attitude object, is likely to be a more valid measure.

---

[34] Dennis Menezes and Norbert F. Elbert, "Alternative Semantic Scaling Formats for Measuring Store Image: An Evaluation," *Journal of Marketing Research,* 16 (February 1979), pp. 80–87.

Assessing the validity of attitude scales is a complex task, and a comprehensive discussion of it is beyond the scope of this textbook.[35] Nevertheless, you should be aware that it includes examining several different types of validity, as outlined next.

Content (Face) Validity *A scale evaluation criterion representing the extent to which the content of a measurement scale appears to tap all relevant facets of the construct it is attempting to measure.*

Construct Validity *A scale evaluation criterion that relates to the following question: "What is the nature of the underlying variable or construct measured by the scale?"*

Convergent Validity *A form of construct validity that represents the extent of association between a measured construct and measures of other constructs with which the measured construct is expected to be related on theoretical grounds.*

Discriminant Validity *A form of construct validity that represents the extent to which a measured construct is not associated with measures of other constructs with which the measured construct is not expected to be related on theoretical grounds.*

Content Validity. Also known as ***face validity,*** the criterion of ***content validity*** represents the extent to which the content of a measurement scale seems to tap all relevant facets of an issue that can influence respondents' attitudes. To illustrate, consider a scale intended for measuring the job satisfaction of industrial salespeople that is made up of multiple items dealing with pay, promotional opportunities, and the quality of supervision. This scale lacks content validity because it lacks items dealing with certain key aspects of a salesperson's job, such as traveling and visiting with customers. Content validity of an attitude scale is a sort of global criterion, but it can only be assessed through a researcher's subjective judgment.

Construct Validity. ***Construct validity*** relates to the question "What is the nature of the underlying variable or construct measured by the scale?" Construct validity of an attitude measure can be assessed quantitatively by computing its correlations with measures of other constructs that one would expect to be strongly associated with the attitude, and measures of constructs that one would *not* expect to be closely tied to the attitude. Strong correlations in the former case are indicative of what is commonly labeled ***convergent validity,*** and weak correlations in the latter case are indicative of ***discriminant validity.*** For a scale to have high construct validity, it must possess both convergent and discriminant validity. A scale purporting to measure attitudes toward abortion can be said to have high construct validity if, for instance, it has relatively strong correlations with measures of attitudes toward respect for life in general (reflecting high convergent validity) and relatively weak correlations with attitudes toward civil rights (reflecting high discriminant validity).

---

[35] Several excellent books and articles on research methods provide comprehensive discussions of validity assessment. See, for example, Claire Selltiz, Lawrence L. Wrightsman, and Stuart W. Cook, *Research Methods in Social Relations* (New York: Holt, Rinehart and Winston, 1976); Nunnally, *Psychometric Theory*; Donald T. Campbell and Donald W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin,* 56 (1959), pp. 81–105; and J. Paul Peter, "Construct Validity: A Review of Basic Issues and Marketing Practices," *Journal of Marketing Research,* 18 (May 1981), pp. 133–145.

Predictive Validity *A scale evaluation criterion representing how well a measured construct is able to predict some other variable or characteristic it is supposed to influence.*

Predictive Validity. As its name implies, ***predictive validity*** deals with the question, "How well does the attitude measure provided by the scale predict some other variable or characteristic it is supposed to influence?" For example, a scale measuring attitudes toward Safeway will have high predictive validity if respondents with relatively high scores on it are found to patronize Safeway more frequently than respondents with relatively low scores on it.

Reliability

Reliability *A criterion for evaluating measurement scales; it represents how consistent or stable the ratings generated by a scale are.*

The ***reliability*** of an attitude scale refers to how consistent or stable the ratings generated by the scale are likely to be. While validity focuses on whether a scale truly measures the construct (and not something else), reliability focuses on whether the scale consistently measures "something" (whatever that "something" may be). Clearly, a good attitude scale must be both valid and reliable. Indeed, unless a scale is reasonably consistent, it cannot be viewed as a true, trustworthy measure of whatever it purports to measure. In this sense, reliability is a necessary, although not sufficient, condition for validity.

Scale reliability can be measured through several different criteria. Two popular criteria are test-retest reliability and split-half reliability.[36]

Test-Retest Reliability *An indicator of the stability of ratings produced by a scale over time.*

Test-Retest Reliability. ***Test-retest reliability*** measures the stability of ratings over time and involves administering the scale to the same group of respondents at two different times. The scale can be considered to have high test-retest reliability if the ratings generated through the two measurements correlate strongly and hence are consistent. The time interval between the two measurements is crucial for test-retest reliability to be a meaningful indicator of a scale's stability. The interval must be long enough so that respondents' ratings during the second measurement are not influenced by their remembering what ratings they gave on the first measurement. But the interval must not be so long that the underlying attitudes change between measurements. Unfortunately, no clear guideline exists for determining an optimum time interval that is sufficiently long yet not too long. A rule of thumb, however, is to allow an interval of about two to four weeks between measurements; attitudes, which are often viewed as being deep-rooted, are unlikely to change during such a short interval.

Split-Half Reliability *An indicator of the degree of consistency across ratings produced by items within a multiple-item scale.*

Split-Half Reliability. ***Split-half reliability*** measures the degree of consistency across items within a scale and can only be assessed for multiple-item scales. It involves splitting the scale items randomly into two sets, with an equal number of items in each set, and examining the correlation between respondents' total scores derived from the two sets of items.

---

[36] Selltiz, Wrightsman, and Cook, *Research Methods in Social Relations,* and Nunnally, *Psychometric Theory,* give excellent, detailed discussions of reliability assessment.

For example, consider a 20-item scale designed to measure customer attitudes toward electronic-banking services. Suppose each item in this scale has five response categories, numbered 1 through 5. After this scale has been administered to a sample of customers, the 20 scale items can be randomly divided into two sets of 10 items. For each respondent the total score in each set will be between 10 (i.e., 10 scale items times the lowest scale value of 1) and 50 (i.e., 10 scale items times the highest scale value of 5). If the 20 scale items are measuring the same underlying construct—namely, customer attitude toward electronic-banking services—a respondent whose total score for the first set of items is, for instance, 38 should have a total score close to 38 for the second set also. Likewise, comparison of the total score for the two sets across respondents should show a consistent pattern of association; the stronger the association, the higher is the split-half reliability of the scale.[37]

Note that the two illustrative methods of assessing a scale's reliability actually focus on two separate dimensions of reliability. Test-retest reliability is a measure of *stability* of the scale items (i.e., the degree to which scores obtained through the scale remain the same *from measurement to measurement over time*), while split-half reliability is a measure of their *equivalency* or *internal consistency* (i.e., the degree to which scores obtained through randomly split halves of the scale correlate with each other *within the same measurement*). Both dimensions are important for a scale's overall reliability.

Ratings on a single-item scale are more susceptible to erratic or random fluctuations (thereby lowering reliability) than summated ratings obtained from a multiple-item scale. Summing the ratings across a number of items has the effect of neutralizing random fluctuations; positive errors in a respondent's ratings on certain items may be offset by negative errors in his or her ratings on other items.[38] Therefore multiple-item scales are generally more reliable than single-item scales.

Sensitivity *A criterion for evaluating measurement scales; it represents the extent to which ratings provided by a scale are able to discriminate between respondents who differ with respect to the construct being measured.*

Sensitivity

The ***sensitivity*** of an attitude scale is closely tied to its reliability, and focuses specifically on its ability to detect subtle differences in the attitudes being measured. A highly sensitive attitude scale should be able to discriminate between respondents who differ even slightly in terms of their attitudes toward something. Alternatively, it should be capable of uncovering minute changes in the same respondent's attitude over time.

---

[37] A more rigorous measure of internal consistency among scale items is called Cronbach's alpha, or coefficient alpha; for details about the procedure involved in computing this measure of reliability, see L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika,* 16 (1951), pp. 297–334. Also see the review article on various aspects of reliability by J. Paul Peter, "Reliability: A Review of Psychometric Basics and Recent Marketing Practices," *Journal of Marketing Research,* 16 (February 1979), pp. 6–17.

[38] Churchill, "A Paradigm for Developing Better Measures."

This description of sensitivity implies that reliability is a prerequisite for it. When an attitude scale is unreliable, one cannot be sure whether differences in attitude scores reflect real differences in attitudes or merely random fluctuations. Therefore, for a scale to be sensitive, it must also be reliable.

Another requirement for a sensitive scale is that it should have a sufficient range of numbers to facilitate detection of fine variations in attitudes (we touched on this point earlier in the chapter). While a single-item scale can theoretically have a large number of rating positions (e.g., questions 1 and 2 in Exhibit 11.3), respondents may not be capable of making such fine judgments. In this regard multiple-item scales have a definite advantage. Even when individual items only have a limited number of response categories, the total scores obtained by summing responses across all items will span a wide range. For example, suppose a multiple-item attitude scale consists of 10 items, each with seven response categories, numbered 1 through 7. Respondents' *total* scores on this 10-item scale can range from 10 to 70. From these total scores respondents can be classified into categories reflecting more subtle variations in attitude than would be reflected if only a single-item, seven-point scale is used.

Our discussion has shown why a multiple-item scale is far superior to a single-item scale when it comes to measuring attitudes. The former is likely to be more valid, reliable, and sensitive than the latter.

SUMMARY

The process of measurement in survey research involves quantifying responses by assigning appropriate numbers to them. However, numbers assigned to survey responses may not possess all the properties commonly attributed to the number system. How closely quantified responses resemble the number system depends on whether they are nominal-, ordinal-, interval-, or ratio-scaled. Nominal-scaled responses have virtually none of the properties of the number system; ratio-scaled responses have all of them; and ordinal- and interval-scaled responses fall in between. Nominal- and ordinal-scaled responses are called nonmetric data because the concept of differences, or distances between numbers, is meaningless for such scales. Interval- and ratio-scaled responses, by virtue of the distance property they possess, are labeled metric data.

Virtually all variables encountered in marketing research surveys can be classified as attributes, behavior, beliefs, or attitudes. Of these four, beliefs and attitudes, whose measurement is the primary focus of this chapter, are relatively more difficult to measure. Scales used to measure beliefs and attitudes are often assumed to have interval properties.

Attitudes have been defined in a number of different ways. But the various definitions seem to agree that attitudes relate to certain inner dispositions of people and that they are multifaceted. These general traits of attitudes have two important implications for their measurement. First, attitudes can only be inferred from indirect measurements. Second, to be effective, an attitude measurement procedure must attempt to tap the various facets capable of influencing overall attitudes.

Five broad, indirect methods for measuring attitudes are available: observing overt behavior, analyzing reactions to partially structured stimuli, evaluating performance on objective tasks, monitoring physiological responses, and using self-report measures. The last method is the most direct of the five approaches and is also the most widely used.

Self-report measures of attitudes typically involve the use of rating scales. A researcher can choose from a wide variety of rating scales, which differ on such dimensions as graphic versus itemized formats, comparative versus noncomparative assessments, forced versus nonforced response choices, balanced versus unbalanced response choices, labeled versus unlabeled response choices, number of scale positions, and measurement level of the data obtained. The choice of an appropriate format for a research project must take into account the nature of the variable to be measured, the ability of respondents to make mental judgments, and the types of analyses to be performed on the collected data.

Three popular multiple-item attitude scales are the Likert, semantic-differential, and Stapel scales. Of the three, the semantic-differential scale is the most widely used, perhaps because of its practical appeal and the visually effective profile diagrams that can be constructed from the data generated. Despite the format differences across the three scales, there are apparently no significant differences in the attitudes inferred from the data generated through them. Therefore a researcher can choose the scale format that best meets the practical requirements of a situation.

To be useful, attitude scales must satisfy three crucial criteria: validity, reliability, and sensitivity. Validity refers to whether a scale truly and fully measures an attitude. It can be assessed in several different ways, including examination of a scale's content or face validity, construct validity (convergent and discriminant validity), and predictive validity. Reliability focuses on the consistency or stability of scores generated by a scale. Two commonly used criteria for reliability assessment are test-retest reliability and split-half reliability. Sensitivity of an attitude scale is the extent to which it is capable of discriminating between respondents with different attitudes. In general, multiple-item scales are better than single-item scales on all three criteria.

KEY TERMS

Measurement
Nominal scale
Central tendency
Mode
Ordinal scale
Median
Interval scale
Mean
Standard deviation
Ratio scale
Nonmetric data
Metric data
Attribute
Behavioral variable
Belief
Attitude
Constitutive (conceptual) definition
Operational definition
Graphic rating scale
Itemized rating scale
Comparative rating scale
Noncomparative rating scale
Forced-choice scale
Nonforced-choice scale
Balanced scale
Unbalanced scale
End piling
Anchor label
Measurement level
Constant-sum scale
Paired-comparison rating scale
Single-item scale
Multiple-item scale
Likert scale
Semantic-differential scale
Stapel scale
Validity
Content (face) validity
Construct validity
Convergent validity
Discriminant validity
Predictive validity
Reliability
Test-retest reliability
Split-half reliability
Sensitivity

QUESTIONS

1. Discuss the advantages and limitations of quantifying survey responses.
2. What are the four key levels of measurement? What mathematical operations are and are not permissible on data from each type of scale?
3. From a survey of users and nonusers of its products, one firm came up with the following inferences:
    a. On the average, users are only half as old as nonusers.
    b. The image that users have of our company is twice as positive as that of nonusers.

    Critically evaluate the meaningfulness and legitimacy of these inferences.
4. Briefly discuss, with concrete examples, whether attitudes are more difficult or less difficult to measure than behavior.
5. What is meant by saying that attitudes are multifaceted? Explain your answer with a specific example.
6. Other than self-report measures, what are the four general methods for inferring attitudes? For each method, state one limitation that you consider to be most serious.
7. "Of all rating scales, graphic rating scales provide the most precise and accurate measures of attitudes." Discuss this statement.

8. When should an itemized rating scale have unbalanced response choices? Why?
9. Examine question 7 in Exhibit 11.3. Are all possible paired comparisons of the five stores listed in it? Can the analysis procedure described in Research in Use 11.1 be applied in its entirety to the data generated by question 7? Why or why not?
10. Suppose a restaurant in your city wants to ascertain the image it has in the minds of its patrons. Construct a five-time scale to measure the perceived image of the restaurant, using each of the multiple-item formats we discussed. Make sure that the five items under each format correspond to the same five dimensions.
11. Which of the three multiple-item scales you constructed in Question 10 would you recommend to the restaurant? Why?
12. Define *scale validity*, *reliability*, and *sensitivity*.
13. Briefly describe the various methods of validity assessment.
14. Can an attitude scale be sensitive if it is not reliable? Why or why not?
15. Explain why a multiple-item scale is a more sensitive measure of attitudes than a single-item scale.

Cases for Part Four

CASE 4.1 *PEOPLE'S NATIONAL BANK*

CASE 4.2 *BANKERS LIFE AND CASUALTY COMPANY*

CASE 4.3 *HERSHEY CHOCOLATE COMPANY (B)*

CASE 4.1

People's National Bank

People's National Bank is located in a city with a population of about 90,000. Its main competitors for consumer accounts in the area are two other banks—First Security Bank and Midtown Bank & Trust—and two savings and loan associations—Prudent Savings Association and Security Savings & Loan. People's National Bank is about the same size (in terms of total assets) as the two competing banks but is somewhat bigger than the two savings and loan associations.

The primary services offered by People's National Bank include checking accounts, savings accounts, mortgage loans, personal or installment loans, trust services, safety deposit boxes, and 24-hour standard financial transactions through automatic-teller machines (ATMs). The other area banks and savings and loan associations also offer these same basic services.

Since the various competing financial institutions in the area offer fairly similar services, the president of People's National Bank, John Goodman, was interested in finding out the extent to which customers were loyal to a given financial institution and how they perceived the five major banks/savings and loan associations in the city. He therefore requested Patricia Stewart, vice president for consumer services, to conduct a mail survey of about 500 current customers of People's National Bank. If this survey were to reveal anything of significance, Ms. Stewart was to conduct a similar survey of noncustomers of People's National Bank to see if similar significant findings emerged.

In his request to Ms. Stewart, Mr. Goodman outlined several specific types of information he would like to see after the survey was conducted and the data analyzed. These items were as follows:

1. What proportion of our bank's current customers use the services of one or more of our main competitors?
2. For which service(s) do our current customers go to our competitors the most?
3. Which of our competitors has the largest proportion of our customers as its customers as well?
4. What is the average number of area financial institutions with which a typical customer of ours has dealings?
5. How do our customers perceive People's National Bank vis-à-vis our competitors along the following dimensions?
   a. Locational convenience,
   b. Convenience of hours of operation.
   c. Friendliness of contact personnel,
   d. Speed of transactions,
   e. Accuracy of transactions,
   f. Appearance of facilities,
   g. Range of services offered,
   h. Technological advancement.
6. How important is each of the above dimensions to our customers?
7. Do customer perceptions of financial institutions along the various dimensions (item 5) and evaluations of the relative importance of those dimensions (item 6) differ according to a customer's sex, age, education, income, and years of association with People's National Bank?

The mail survey questionnaire was to be sent to a random sample of 500 names and addresses selected from the current customer base of People's National Bank. A business reply envelope would also be sent for returning the questionnaires to the bank.

QUESTIONS

1. Design a suitable mail questionnaire (including a cover letter and all necessary instructions) that can meet the information needs of Mr. Goodman.
2. On the basis of Mr. Goodman's overall objective implied in the list, is the proposed mail survey the most appropriate first step? If not, what else would you propose?
3. What are the pros and cons of limiting the proposed survey to just People's National Bank's customers?

CASE 4.2

Bankers Life & Casualty Company

The marketing research department of Bankers Life and Casualty Company, headquartered in Northbrook, Illinois, conducted a mail survey of new buyers of the company's auto and homeowners/renters insurance. The questionnaire used in this survey was in the form of a booklet that was professionally prepared. The first page of this booklet (i.e., the outside of the front cover) contained the cover letter, which is reproduced here as Exhibit 1.

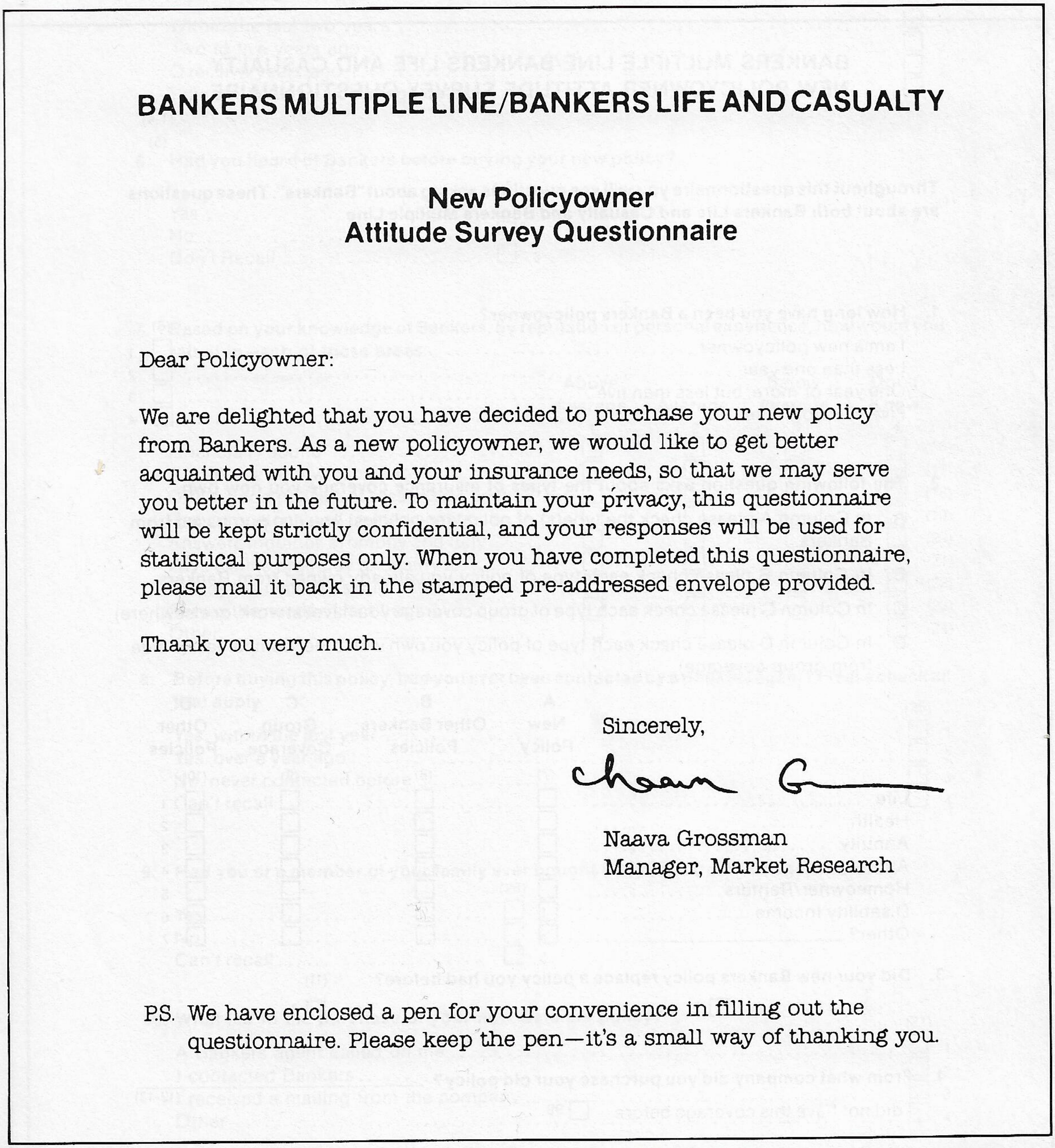
The questionnaire began on the inside front cover and was 10 pages long. These 10 pages are reproduced (in condensed form) as Exhibit 2. (*Note:* A sample page from this questionnaire appeared in Chapter 6.)

QUESTIONS

1. What are the strengths and weaknesses of the cover letter shown in Exhibit 1?

2. Critically evaluate the mail questionnaire (Exhibit 2) on dimensions such as format, wording, ease of answering, and question sequencing.

Exhibit 1 *Bankers Cover Letter*



BANKERS MULTIPLE LINE/BANKERS LIFE AND CASUALTY

New Policyowner
Attitude Survey Questionnaire

Dear Policyowner:

We are delighted that you have decided to purchase your new policy from Bankers. As a new policyowner, we would like to get better acquainted with you and your insurance needs, so that we may serve you better in the future. To maintain your privacy, this questionnaire will be kept strictly confidential, and your responses will be used for statistical purposes only. When you have completed this questionnaire, please mail it back in the stamped pre-addressed envelope provided.

Thank you very much.

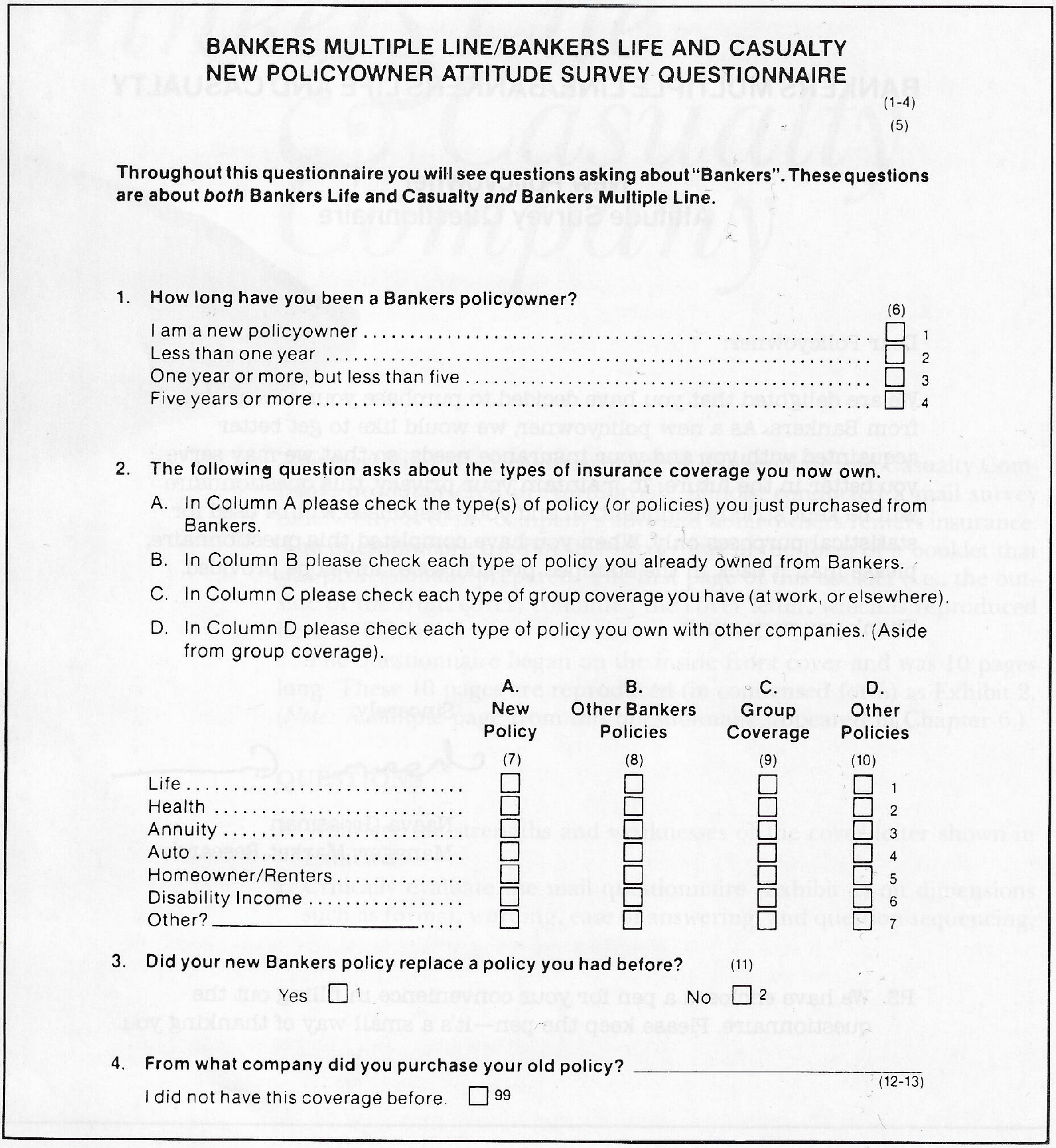
Sincerely,

Naava Grossman
Manager, Market Research

P.S. We have enclosed a pen for your convenience in filling out the questionnaire. Please keep the pen—it's a small way of thanking you.

Courtesy of Marketing Research Department, Bankers Life and Casualty Company, 1000 Sunset Ridge Road, Northbrook, IL 60062.

Exhibit 2 *Bankers Questionnaire*



BANKERS MULTIPLE LINE/BANKERS LIFE AND CASUALTY
NEW POLICYOWNER ATTITUDE SURVEY QUESTIONNAIRE

(1-4)
(5)

Throughout this questionnaire you will see questions asking about "Bankers". These questions are about *both* Bankers Life and Casualty *and* Bankers Multiple Line.

1. **How long have you been a Bankers policyowner?**

	(6)
I am a new policyowner	☐ 1
Less than one year	☐ 2
One year or more, but less than five	☐ 3
Five years or more	☐ 4

2. **The following question asks about the types of insurance coverage you now own.**
   A. In Column A please check the type(s) of policy (or policies) you just purchased from Bankers.
   B. In Column B please check each type of policy you already owned from Bankers.
   C. In Column C please check each type of group coverage you have (at work, or elsewhere).
   D. In Column D please check each type of policy you own with other companies. (Aside from group coverage).

	A. New Policy (7)	**B. Other Bankers Policies** (8)	**C. Group Coverage** (9)	**D. Other Policies** (10)	
Life	☐	☐	☐	☐	1
Health	☐	☐	☐	☐	2
Annuity	☐	☐	☐	☐	3
Auto	☐	☐	☐	☐	4
Homeowner/Renters	☐	☐	☐	☐	5
Disability Income	☐	☐	☐	☐	6
Other? ______________	☐	☐	☐	☐	7

3. **Did your new Bankers policy replace a policy you had before?** (11)

   Yes ☐ 1 No ☐ 2

4. **From what company did you purchase your old policy?** ______________________ (12-13)

   I did not have this coverage before. ☐ 99

(*continued*)

5. **How long ago did you purchase your old policy?** (14)

Within the last two years .... ☐ 1
Two to five years ago .... ☐ 2
Over five years ago .... ☐ 3
I did not have this coverage before .... ☐ 4

6. **Had you heard of Bankers before buying your new policy?** (15)

Yes .... ☐ 1
No .... ☐ 2
Don't Recall .... ☐ 3

7. **Based on your knowledge of Bankers, by reputation or personal experience, how would you rate it in each of these areas:**

	Above Average 1	Average 2	Below Average 3	Don't Know 4	
Financially sound	☐	☐	☐	☐	(16)
A large, well known company	☐	☐	☐	☐	(17)
Offers good policies at reasonable cost ...	☐	☐	☐	☐	(18)
Has competent, well trained agents	☐	☐	☐	☐	(19)
Answers inquiries promptly and fully	☐	☐	☐	☐	(20)
Pays claims promptly and fairly	☐	☐	☐	☐	(21)
Serves the needs of senior citizens	☐	☐	☐	☐	(22)
Serves the needs of families	☐	☐	☐	☐	(23)
Other ________	☐	☐	☐	☐	(24)

8. **Before buying this policy, had you ever been contacted by a Bankers agent? Please check all that apply.** (25)

Yes, within the last year .... ☐ 1
Yes, over a year ago .... ☐ 2
No, never contacted before .... ☐ 3
Can't recall .... ☐ 4

9. **Had you or a member of your family ever bought insurance from this agent before?** (26)

Yes .... ☐
No .... ☐ 2
Can't recall .... ☐ 3

10. **What led to the purchase of your new Bankers policy?** (27)

A Bankers agent called on me .... ☐ 1
I contacted Bankers .... ☐ 2
I received a mailing from the company .... ☐ 3
Other .... ☐ 4

(continued)

11. How would you rate the Bankers agent in each of these areas?

	Excellent 1	**Good** 2	**Fair** 3	**Poor** 4	
A. Being courteous and businesslike	☐	☐	☐	☐	(28)
B. Having a thorough knowledge of insurance	☐	☐	☐	☐	(29)
C. Describing policies, costs, etc., clearly and openly	☐	☐	☐	☐	(30)
D. Helping you select the right coverage	☐	☐	☐	☐	(31)
E. Following through to get your policy issued to you smoothly and quickly ...	☐	☐	☐	☐	(32)

12. Based on your experience with the agent who sold you this policy, would you buy insurance from him or her again?

(33)

Yes .... ☐ 1
No .... ☐ 2
Maybe .... ☐ 3

13. Which of the following statements were true in your case when you decided to buy your Bankers policy?

	True in my case 1	**Not true in my case (or does not apply)** 2	
A. My Bankers policy gives me better coverage than my old policy..................	☐	☐	(34)
B. The Bankers agent did a better job than the agent of my old company..............	☐	☐	(35)
C. A friend or relative recommended Bankers.	☐	☐	(36)
D. I have had good experience with Bankers.	☐	☐	(37)
E. My old company just raised the rates.	☐	☐	(38)
F. I now have all my insurance with Bankers.	☐	☐	(39)
G. I have had bad experience with my old company.	☐	☐	(40)
H. I saved money with Bankers.	☐	☐	(41)

I. Other important reasons why I decided to purchase from Bankers were ____________ (42)

____

14. Besides Bankers and your previous company, did you get a quote from any other insurance company?

(43)

Yes .... ☐ 1
No .... ☐ 2

(*continued*)

15. Below are examples of services that a person might receive after becoming a Bankers policyholder. Please indicate how interested you are in receiving that kind of service.

	Very Interested	Somewhat Interested	Not Interested	
	1	2	3	
A. Reviews to determine whether your coverages are adequate for your current needs.	☐	☐	☐	(45)
B. Reviews to make sure your policy information is up to date (addresses, drivers, home additions, etc.)	☐	☐	☐	(46)
C. Information about additional options or other insurance products offered by Bankers.	☐	☐	☐	(47)

16. Would you ever be interested in discussing any of the following with a Bankers agent? (Check all that apply)

(48)

Life Insurance ☐ 1
Health Insurance ☐ 2
Disability Income ☐ 3
Annuities ☐ 4
Auto Insurance ☐ 5
Homeowners/Renters ☐ 6
Other ________________ ☐ 7
No interest ☐ 8

17. How often would you like to have someone from Bankers get in touch with you to review your insurance?

(49)

At least once a year ☐ 1
Once every year or two ☐ 2
Only at my request ☐ 3

18. Should the contact be by: (Check all that apply)

(50)

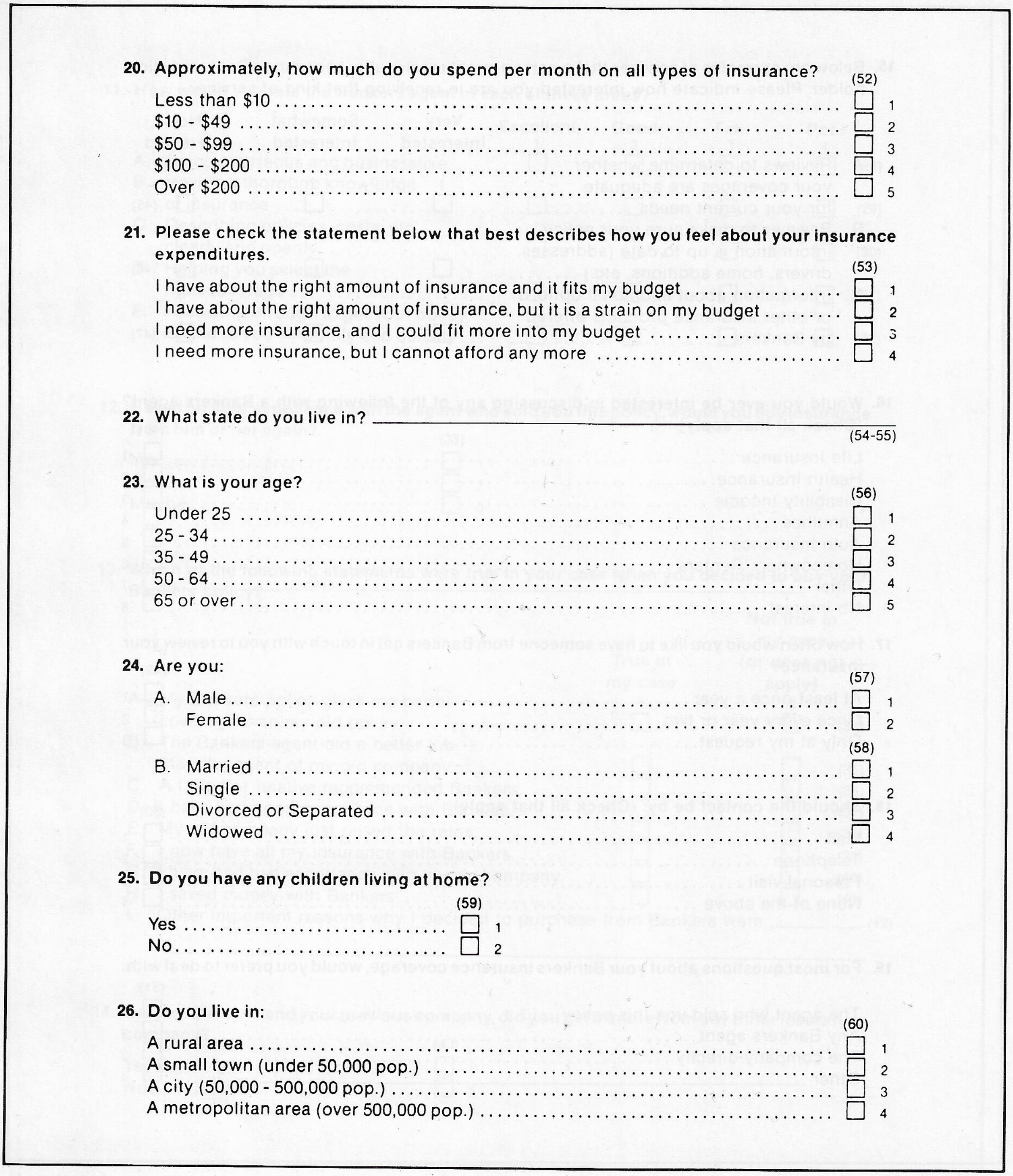
Mail ☐ 1
Telephone ☐ 2
Personal visit ☐ 3
None of the above ☐ 4

19. For most questions about your Bankers insurance coverage, would you prefer to deal with:

(51)

The agent who sold you this policy ☐ 1
Any Bankers agent ☐ 2
The company directly ☐ 3
Other ________________ ☐ 4

(continued)



20. Approximately, how much do you spend per month on all types of insurance?

(52)

Less than $10 ☐ 1
$10 - $49 ☐ 2
$50 - $99 ☐ 3
$100 - $200 ☐ 4
Over $200 ☐ 5

21. Please check the statement below that best describes how you feel about your insurance expenditures.

(53)

I have about the right amount of insurance and it fits my budget ☐ 1
I have about the right amount of insurance, but it is a strain on my budget ☐ 2
I need more insurance, and I could fit more into my budget ☐ 3
I need more insurance, but I cannot afford any more ☐ 4

22. What state do you live in? ____________________

(54-55)

23. What is your age?

(56)

Under 25 ☐ 1
25 - 34 ☐ 2
35 - 49 ☐ 3
50 - 64 ☐ 4
65 or over ☐ 5

24. Are you:

(57)

A. Male ☐ 1
Female ☐ 2

(58)

B. Married ☐ 1
Single ☐ 2
Divorced or Separated ☐ 3
Widowed ☐ 4

25. Do you have any children living at home?

(59)

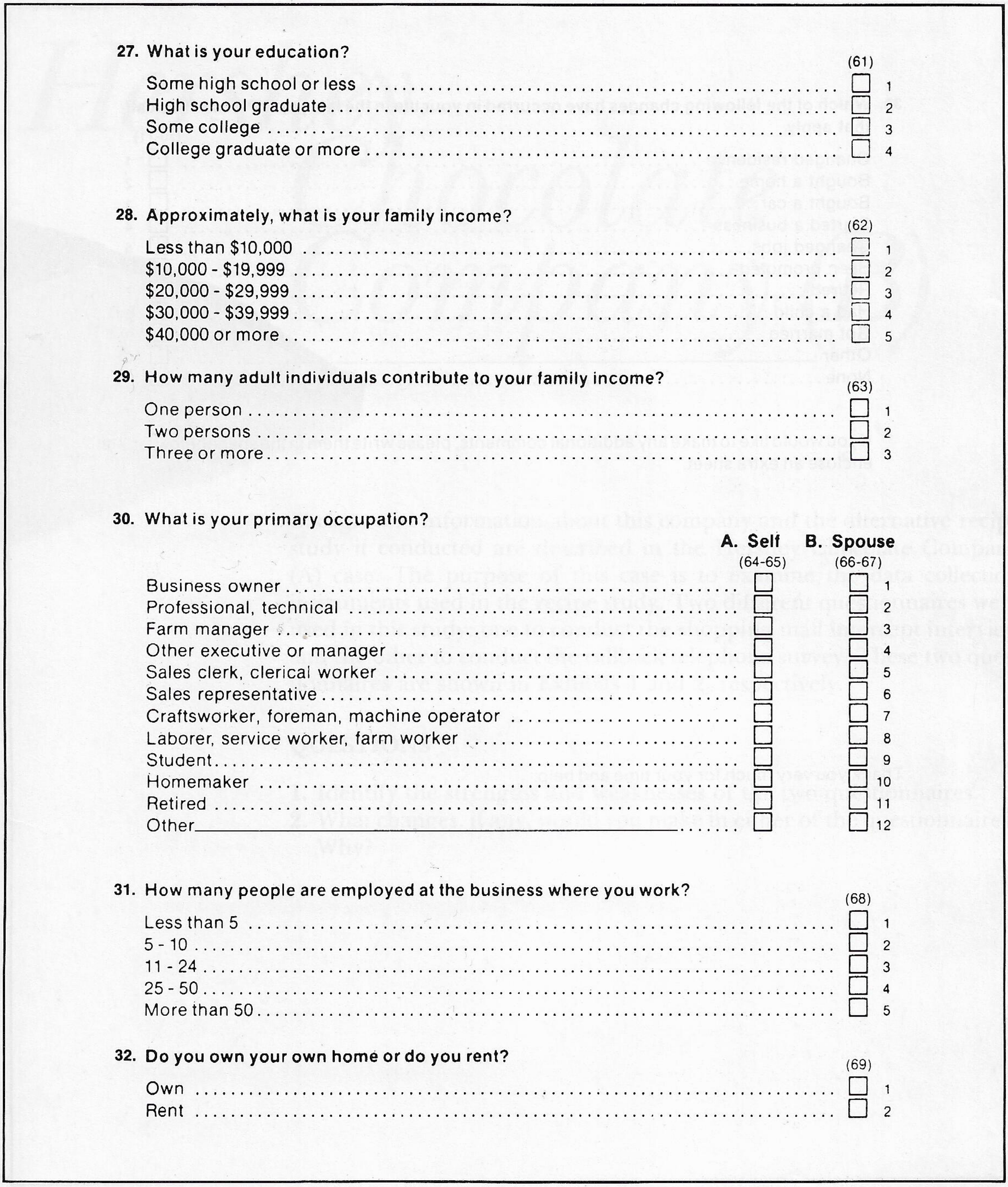
Yes ☐ 1
No ☐ 2

26. Do you live in:

(60)

A rural area ☐ 1
A small town (under 50,000 pop.) ☐ 2
A city (50,000 - 500,000 pop.) ☐ 3
A metropolitan area (over 500,000 pop.) ☐ 4

(continued)



27. What is your education?

	(61)	
Some high school or less	☐	1
High school graduate	☐	2
Some college	☐	3
College graduate or more	☐	4

28. Approximately, what is your family income?

	(62)	
Less than $10,000	☐	1
$10,000 - $19,999	☐	2
$20,000 - $29,999	☐	3
$30,000 - $39,999	☐	4
$40,000 or more	☐	5

29. How many adult individuals contribute to your family income?

	(63)	
One person	☐	1
Two persons	☐	2
Three or more	☐	3

30. What is your primary occupation?

	A. Self (64-65)	B. Spouse (66-67)	
Business owner	☐	☐	1
Professional, technical	☐	☐	2
Farm manager	☐	☐	3
Other executive or manager	☐	☐	4
Sales clerk, clerical worker	☐	☐	5
Sales representative	☐	☐	6
Craftsworker, foreman, machine operator	☐	☐	7
Laborer, service worker, farm worker	☐	☐	8
Student	☐	☐	9
Homemaker	☐	☐	10
Retired	☐	☐	11
Other________________	☐	☐	12

31. How many people are employed at the business where you work?

	(68)	
Less than 5	☐	1
5 - 10	☐	2
11 - 24	☐	3
25 - 50	☐	4
More than 50	☐	5

32. Do you own your own home or do you rent?

	(69)	
Own	☐	1
Rent	☐	2

(*continued*)

33. Which of the following changes have occurred in your life in the last year? Please check all that apply.

		(70-71)
Changed residence	☐	1
Bought a home	☐	2
Bought a car	☐	3
Started a business	☐	4
Changed jobs	☐	5
Been promoted	☐	6
Retired	☐	7
Had a child	☐	8
Got married	☐	9
Other ____________	☐	10
None	☐	11

If you would like to make any additional comments, please write them in the space below, or enclose an extra sheet. (72)

Thank you very much for your time and help.

Courtesy of Marketing Research Department, Bankers Life and Casualty Company, 1000 Sunset Ridge Road, Northbrook, IL 60062.

CASE 4.3

Hershey Chocolate Company (B)

Background information about this company and the alternative recipe study it conducted are described in the Hershey Chocolate Company (A) case. The purpose of this case is to examine the data collection instruments used in the recipe study. Two different questionnaires were used in this study: one to conduct the shopping mall intercept interview and the other to conduct the callback telephone survey. These two questionnaires are shown in Exhibits 1 and 2, respectively.

QUESTIONS

1. Identify the strengths and weaknesses of the two questionnaires.
2. What changes, if any, would you make in either of the questionnaires? Why?

Exhibit 1 *Recipe Study, Mall Placement*

TIME STARTED: _____:_____ ( ) AM ( ) PM

Hello, I'm ______________ from Interviewing Services. We're making a study with women today.

1a. First, are you the female head of your household?
1( ) Yes 2( ) No (TERMINATE AND TALLY)

1b. Are you between the ages of 18 and 49?
1( ) Yes 2( ) No (TERMINATE AND TALLY)

1c. How many children, if any, under 17 years of age do you have living at home?
# OF CHILDREN: ______________________ -9
(IF NONE, TERMINATE AND TALLY)

1d. Do you or anyone else in your household work for an advertising agency, a marketing research company or a manufacturer, distributor or retailer of candy?
1( ) Yes (TERMINATE AND TALLY) 2( ) No

1e. Which, if any, of the following products have you used in the past few months? (READ LIST)

Pillsbury flour	1( ) Yes	2( ) No	-10
Land O'Lakes Butter	1( ) Yes	2( ) No	-11
Peanut Butter Chips	1( ) Yes	2( ) No	-12

IF YES TO PEANUT BUTTER CHIPS, RESPONDENT QUALIFIES FOR USER QUOTA.
ASK Q. 2 NEXT.

1f. IF NOT PEANUT BUTTER CHIPS USER, ASK:
Have you done any baking from scratch in the past few months?

13-

	Yes	1
(TERMINATE AND TALLY)	No	2
	N/A	9

1g. Does your family like baked peanut butter flavored desserts or snacks?

14-

	Yes	1
(TERMINATE AND TALLY)	No	2
	N/A	9

2. CHECK QUOTA:

15-

Peanut Butter Chips user	1
Scratch baker	2

3a. HAND RESPONDENT BOARD #231.
Please look at this picture and read the recipe written beside it. Take all the time you want to study the picture and recipe.
ALLOW RESPONDENT TIME.
How much does this appeal to you in terms of being appetizing? POINT TO DESSERT/SNACK IN THE PICTURE.
Would you say it is:

(continued)

16-

Very appealing	5
Somewhat appealing,	4
Neither appealing nor unappealing,	3
Somewhat unappealing, or	2
Very unappealing?	1

3b. Why is it (NAME ANSWER TO #3a) to you? (PROBE)

_______________ 17-18 __ __

_______________ 19-20 __ __

_______________ 21-22 __ __

_______________ 23-24 __ __

_______________ 25-26 __ __

_______________ 27-28 __ __

3c. If you were reading a magazine and saw this recipe, how likely do you think you would be to make it? Would you be:

29-

Very likely to make it,	5
Somewhat likely to make it,	4
Neither likely nor unlikely to make it,	3
Somewhat unlikely to make it, or	2
Very unlikely to make it?	1

TAKE BACK BOARD.

4a. HAND RESPONDENT BOARD #479.

Please look at this picture and read the recipe written beside it. Take all the time you want to study the picture and recipe.

ALLOW RESPONDENT TIME.

How much does this appeal to you in terms of being appetizing? POINT TO DESSERT/SNACK IN THE PICTURE. Would you say it is:

30-

Very appealing,	5
Somewhat appealing,	4
Neither appealing nor unappealing,	3
Somewhat unappealing, or	2
Very unappealing?	1

(continued)

4b. Why is it (NAME ANSWER TO #4a) to you? (PROBE)

______________________________ 31-32 __ __

______________________________ 33-34 __ __

______________________________ 35-36 __ __

______________________________ 37-38 __ __

______________________________ 39-40 __ __

41-42 __ __

4c. If you were reading a magazine and saw this recipe, how likely do you think you would be to make it? Would you be:

43-

Very likely to make it,	5
Somewhat likely to make it,	4
Neither likely nor unlikely to make it,	3
Somewhat unlikely to make it, or	2
Very unlikely to make it?	1

TAKE BACK BOARD.

5a. HAND RESPONDENT BOARD #658.

Please look at this picture and read the recipe written beside it. Take all the time you want to study the picture and recipe.

ALLOW RESPONDENT TIME.

How much does this appeal to you in terms of being appetizing? POINT TO DESSERT/SNACK IN THE PICTURE. Would you say it is:

44-

Very appealing,	5
Somewhat appealing,	4
Neither appealing nor unappealing,	3
Somewhat unappealing, or	2
Very unappealing?	1

5b. Why is it (NAME ANSWER TO #5a) to you? (PROBE)

______________________________ 45-46 __ __

______________________________ 47-48 __ __

______________________________ 49-50 __ __

______________________________ 51-52 __ __

______________________________ 53-54 __ __

55-56 __ __

(*continued*)

5c. If you were reading a magazine and saw this recipe, how likely do you think you would be to make it? Would you be:

57-

Very likely to make it,	5
Somewhat likely to make it,	4
Neither likely nor unlikely to make it,	3
Somewhat unlikely to make it, or	2
Very unlikely to make it?	1

TAKE BACK BOARD.

6a. SHOW RESPONDENT ALL 3 BOARDS.
Now that you have seen all three recipes, which one would you be most likely to use?

58-

#231 (Semi-Sweet Chocolate Chips, regular cookie)	1
#479 (Semi-Sweet Chips, bar cookie)	2
#658 (Syrup, brownie)	3

6b. Why would you be most likely to make that one? (PROBE)

______ 59-60 __ __

______ 61-62 __ __

______ 63-64 __ __

______ 65-66 __ __

______ 67-68 __ __

69-70 __ __

7a. I would like to give you a copy of this recipe, Peanut Butter Chips and (NAME PRODUCT CHOSEN IN Q. 6a) so that you can make this recipe at home. In addition for your cooperation, we will give you $5.00 and the two products.

We would like you to make the recipe in the next few days and we will call you to see how much you liked the recipe.

1( ) Will cooperate 2( ) Will not cooperate (TERMINATE AND TALLY)

7b. When would be the best day and time to call you within the next 3 or 4 days?

DAY: ______ TIME: ______

MUST BE WITHIN NEXT 4 DAYS.

7c. What is your telephone number? ______

MUST BE GIVEN TO QUALIFY.

8a. And a few final questions for classification purposes. Including yourself, how many people are there in your household?

IN HOUSEHOLD: ______ 71-72

(*continued*)

8b. Into which of the following age groups do you fall? (READ LIST)

73-

18 to 19	1
20 to 29	2
30 to 39	3
40 to 49	4

8c. And into which of the following groups does your total household income fall? HAND CARD TO RESPONDENT. Just read me the appropriate letter.

74-

"A" Under $10,000	1
"B" $10,000 to $14,999	2
"C" $15,000 to $19,999	3
"D" $20,000 to $24,999	4
"E" $25,000 to $29,999	5
"F" $30,000 or more	6

9a. GIVE RESPONDENT RECIPE, PRODUCTS AND $5.00.
Here is the recipe, two of the ingredients, and $5.00. We will call you on (NAME ANSWER TO Q. 7b) to get your reactions to the recipe.

9b. CHECK RECIPE AND PRODUCTS PLACED:

75-

#231 (Peanut Butter Chips and Semi-Sweet Chocolate Chips, regular cookie)	1
#479 (Peanut Butter Chips and Semi-Sweet Chips, bar cookie)	2
#658 (Peanut Butter Chips and Syrup, brownies)	3

TIME ENDED: _____:_____ ( ) AM ( ) PM 76-77 __ __

RESPONDENT ____________________

ADDRESS ____________________

CITY ____________ STATE ____________ ZIP ________

INTERVIEWER ____________________ 78-79

DATE ____________________ 80-1

Courtesy of Hershey Foods Corporation.

Exhibit 2 *Recipe Study Callback*

TIME START: _____:_____ ( ) AM ( ) PM

CHECK RECIPE:

4-

#231 (Peanut Butter Chips and Semi-Sweet Chocolate Chips, regular cookie)	1
#479 (Peanut Butter Chips and Semi-Sweet Chips, bar cookie)	2
#658 (Peanut Butter Chips and syrup, brownies)	3

Hello, I'm ____________ from Interviewing Services. May I please speak with (NAME RESPONDENT)?.

1a. First, did you make the recipe using the ingredients we gave you at the Mall the other day?

5-

(#2 NEXT)	Yes	1
	No	2

1b. IF NO:
Are you planning to make the recipe in the next few days?

6-

	Yes	1
(TERMINATE AND TALLY)	No	2

1c. IF YES:
When can I call you to get your reactions to the recipe?
DAY: ______________________ TIME: ______________________

2a. How easy was the recipe for you to make—would you say it was:

7-

(#3a NEXT)	Very easy to make,	5
	Somewhat easy to make,	4
	Neither easy nor difficult,	3
	Somewhat difficult to make, or	2
	Very difficult to make?	1

(continued)

2b. IF SOMEWHAT/VERY DIFFICULT:
What did you find difficult about making it? (PROBE)

______________________________ 8- 9 __ __
______________________________ 10-11 __ __
______________________________ 12-13 __ __
______________________________ 14-15 __ __
______________________________ 16-17 __ __

3a. Did the finished product taste:

18-

	Better than you expected,	3
(#3c NEXT)	About the same as you expected, or	2
	Not as good as you expected?	1

3b. IF BETTER/NOT AS GOOD:
In what way was it (NAME ANSWER TO #3a)?

______________________________ 19-20 __ __
______________________________ 21-22 __ __
______________________________ 23-24 __ __
______________________________ 25-26 __ __
______________________________ 27-28 __ __
29-30 __ __

3c. Did you serve this as:

A snack for yourself and family,	01	31-32 __ __
Dessert for yourself and family,	02	33-34 __ __
A snack for guests,	03	35-36 __ __
Dessert for guests,	04	
A special occasion, or (SPECIFY) ______________________________		
Something else? (SPECIFY) ______________________________		

(*continued*)

4a. How likely do you think you will be to make this recipe again. Would you say you will:

37-

(#4c NEXT)	Definitely make it again,	5
	Probably make it again,	4
GO TO RESPONDENT INFORMATION AT BOTTOM OF PAGE	May or may not make it again,	3
	Probably not make it again, or	2
	Definitely not make it again?	1

4b. IF PROBABLY/DEFINITELY NOT MAKE AGAIN:
Why do you think you won't make it again? (PROBE)

______________________ 38-39 __ __
______________________ 40-41 __ __
______________________ 42-43 __ __
______________________ 44-45 __ __
______________________ 46-47 __ __

(GO TO RESPONDENT INFORMATION AT BOTTOM OF PAGE)

4c. IF DEFINITELY/PROBABLY MAKE AGAIN:
What was the one thing you liked most about this recipe?

______________________ 48-49 __ __
______________________ 50-51 __ __
______________________ 52-53 __ __
______________________ 54-55 __ __
______________________ 56-57 __ __
58-59 __ __

TIME ENDED: _____ : _____ ( ) AM ( ) PM 60-61 __ __

RESPONDENT ______________________ PHONE ____________
ADDRESS ______________________
CITY ____________ STATE ____________ ZIP ____________
INTERVIEWER ______________________ 62-63
DATE ____________ 64-79 (BLANK)
80-2

STAPLE TO BACK OF APPROPRIATE PLACEMENT INTERVIEW.

Courtesy of Hershey Foods Corporation.

PART FIVE

DESIGNING the Sample & COLLECTING the Data

CHAPTER 12 *Introduction to Sampling*

CHAPTER 13 *Advanced Probability-Sampling Methods*

CHAPTER 14 *Nonprobability-Sampling Methods*

CHAPTER 15 *Data Collection: Fieldwork and Potential Errors*

CASES FOR PART FIVE

CHAPTER 12

Introduction to Sampling

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Define and distinguish between sampling and census studies.*
- *Define the terms population and sampling frame.*
- *Explain when census studies should be done.*
- *State the difference between probability and nonprobability sampling.*
- *Define simple random sampling and describe how a simple random sample is chosen.*
- *Distinguish between a parameter and a statistic.*
- *Explain what is meant by a sampling distribution.*
- *Construct confidence intervals for population means and proportions.*
- *Enumerate the factors to be considered in determining sample size.*
- *Compute the required sample size to achieve a certain degree of precision at a desired confidence level.*
- *Explain the role of incidence rate of a characteristic in determining sample size.*
- *List the limitations of simple random sampling.*

WHAT IS SAMPLING

Consider the following items.

Item. The ABC News–Harris Survey is a study that is done three times a week in order to constantly monitor the reactions of the American public to a wide variety of economic, political, social, and other issues. The importance and popularity of this survey is evident from its being widely cited in newspapers as well as on radio and TV. Each ABC News–Harris Survey is based on responses from 1600 adults spread over 200 locations across the country.[1]

Item. The A. C. Nielsen Company's Nielsen Retail Index offers valuable sales and brand share data on a regular basis to manufacturers of a wide variety of consumer products such as food, drugs, and cosmetics. These sales and brand share estimates are based on a representative sample of 1600 stores scattered across the country.[2]

One important feature common to most research studies emerges from these items: Insights and inferences that are crucial to decision makers can be based on data from a *sample* of units, which is quite small in size compared with the size of the total group of units of interest to the decision makers. What is termed as *sampling* is the starting point of the inferential process involved in marketing research studies.

---

[1] "Survey Methodology," a brochure prepared by Louis Harris and Associates, 630 Fifth Avenue, New York, NY 10020.

[2] "Management with the Nielsen Retail Index," a brochure prepared by the A. C. Nielsen Company, Nielsen Plaza, Northbrook, IL 60062.

Definition. ***Sampling*** is the selection of a fraction of the total amount of units of interest to decision makers, for the ultimate purpose of being able to draw general conclusions about the entire body of units.

Definition. The entire body of units of interest to decision makers in a situation is known as the ***population,*** or ***universe.***

Definition. A listing of population units from which a sample is chosen is known as a ***sampling frame.***

To clarify the distinction between *population* and *sampling frame,* let us consider the following scenario. A university offering a graduate program in business wants to conduct a study to determine the criteria prospective students use in evaluating different graduate business programs. In this scenario, the population can be defined as "all individuals with an interest in pursuing a graduate degree in business within the next year." How can one identify such individuals? One approach is to contact the organization that administers the Graduate Management Admissions Test (GMAT) and obtain from it a list of individuals who took the test during the past twelve months. Such a list constitutes a sampling frame. Note that this sampling frame will contain most, *but not all,* of the individuals implied in the definition of the population (e.g., some individuals who did not take the GMAT during the past twelve months, and therefore excluded from the sampling frame, may nevertheless be interested in pursuing a graduate business degree). However, since generating an exhaustive list of population units is often impractical or prohibitively expensive, sampling frames that are more readily obtainable and are good approximations of the population are frequently used to select samples.

In marketing research studies the meaning of the term *population* is not restricted to people per se. Rather, it is used to denote any group of units–human or nonhuman–about which researchers or decision makers want to make some inferences. For instance, the implied population in the studies conducted to determine the Nielsen Retail Index is a collection of nonhuman units–all grocery, drug, and similar stores.

Census Study *A study involving all units in a population.*

The basic reason sampling is an important component of most marketing research projects is that a study based on a sample of units has distinct advantages over a ***census study,*** which involves obtaining data from every unit in a population. For instance, a census study would be involved in determining the Nielsen Retail Index if data for the index were to be collected from *all* stores selling the categories of products to which the index pertains. Let us now examine the advantages of a sampling study over a census study.

SAMPLING VERSUS CENSUS STUDIES

One of the main advantages of a sampling study is its lower cost relative to a census study. A key factor that directly affects the data collection and data analysis costs in any research project is the number of units studied. Irrespective of whether the data collection method in a research project involves a personal-interview survey, a telephone survey, a mail survey, or an observation, the larger the number of units from which data are obtained, the greater are the data collection costs and the data analysis costs (because of the larger volume of data that have to be analyzed). In most real-life research projects the population sizes are prohibitively large, and the money allocated for research is too limited for conducting census studies.

A second advantage of a sampling study over a census study is in terms of the time needed for a research project. Given a set of limited resources for conducting research (e.g., money, fieldworkers to collect the data, facilities to code and analyze the data), the research will take longer to complete if it involves a census study rather than a sampling study. Data that are not obtained on a timely basis may not be relevant for decision making; worse still, the use of such data may result in erroneous decisions.

By now you are probably wondering: "Well, a census study is more expensive and time-consuming than a sampling study; but isn't it more *accurate* than a sampling study, since it obtains data from each and every unit within a population?" Surprisingly, the answer to that question is, "Not necessarily." By its definition, a census study involves looking at every unit in a population. However, the accuracy of data obtained from a study depends not only on the number of population units included in it but also on a host of other factors related to the process of collecting the data.

The *total error* associated with a research project may contain one or both of the following:

1. *Sampling error* that will occur in any project involving a sampling study, merely because only a sample of the population is studied (sampling error will be discussed in detail later in this chapter).
2. Other *nonsampling errors* that may occur during the process of data collection and analysis.

Nonsampling errors can arise from a multitude of factors, such as poor questionnaire construction, ill-trained fieldworkers, errors on the part of respondents, and errors in coding responses. The best way for a researcher to minimize nonsampling errors is to have adequate control over the entire process of data gathering, coding, and analysis—for example, by using properly trained fieldworkers, by giving them proper instructions, and by closely supervising their work. A sampling study, by virtue of its limited scope when compared with a census study, is much more likely to offer a researcher control over nonsampling errors. We will discuss nonsampling errors in greater detail in Chapter 15.

Let us look at an example to illustrate this point. Suppose you want to do a census study of "all homes in the United States that have video recorders," using personal interviews to gather the data that you need. A population like this one will contain at least a few million units. Consequently, even if you have enough time and money to conduct a census study, you will probably have difficulty identifying enough *qualified* interviewers, let alone recruiting them, training them, and supervising them adequately. A researcher planning to conduct a census study under such circumstances will have to use marginally qualified personnel and settle for reduced control over the research. Huge nonsampling errors can result, which will more than offset the absence of any sampling error.

Therefore in most cases a well-conducted sampling study, in spite of some unavoidable sampling error, is likely to have a lower total error than a census study. This conclusion is supported by a somewhat surprising activity of the U.S. Census Bureau; the bureau actually conducts some controlled sampling studies *to verify the accuracy* of the data generated by its decennial population census studies.

In addition to the cost, time, and accuracy advantages that we have discussed so far, sampling studies have another indirect benefit over census studies in situations involving human populations. Because of the increasing use of research studies by a variety of institutions, people are becoming more likely to be requested to participate in surveys. Consequently, fresh respondents—that is, those who have not participated in any research surveys—are becoming harder to find than ever before. Indeed, a recent survey conducted by Walker Research, Inc., discovered that 70% of the United States adult population had participated in some research project or another, and that 33% of the United States adult population had taken part in some survey within the past year.[3]

Using fresh respondents in research projects is beneficial for two reasons:

1. A fresh respondent is more likely to agree to cooperate with a researcher and provide the needed data than one who is tired of participating in research surveys.
2. A respondent who has participated in several research surveys may, in effect, be a professional respondent. That is, such a respondent may unintentionally, or even deliberately, give biased responses, which in turn can add to nonsampling errors.

What does this discussion of fresh respondents have to do with our discussion of sampling versus census studies? Census studies increase the chances of multiple survey participation by respondents in many human populations that are probably already being oversampled. Hence in the interest of conserving the available pool of fresh respondents in the long run, a sampling study is preferable over a census study.

---

[3] "Results of Walker's 1986 Study Indicate Favorable Industry Image," *The Marketing Researcher* 16 (3), p. 2 (Walker Research, Inc., 3939 Priority Way South Drive, Indianapolis, IN 46280); Stephen W. McDaniel, Perry Verille, and Charles S. Madden, "The Threats to Marketing Research: An Empirical Reappraisal," *Journal of Marketing Research,* 22 (February 1985), pp. 74–80.

CONDITIONS FOR DOING CENSUS STUDIES IN MARKETING

Though census studies have several important limitations, we do not mean to say that they should never be conducted in marketing situations. Notice that the cost, time, and accuracy drawbacks of a census study will not be significant *if the population is small.* For example, populations such as "all neurosurgeons in the Chicago metropolitan area," "all savings and loan associations in Arkansas" and "all manufacturers who have a need for gold in its raw form" are unlikely to contain a large number of units. In research projects involving small populations, a census is certainly feasible from a cost, time, and accuracy standpoint. However, the feasibility of a census study in a situation does not necessarily imply that it must be conducted, as the next example shows.

Example. Consider a hypothetical population consisting of 200 banks in a certain region. Let us say we want to determine the average number of tellers employed per bank in this population of banks. Now a census study of the banks will be feasible, given the relatively small size of the population. However, suppose we know that all the 200 banks are *identical* in terms of the number of tellers each uses, although we do not know what that number is. Given this knowledge about the nature of the population of banks, should we do a census study? We would be wasting our resources if we did. Since we know that all the banks are identical, all we need to do is study a sample of just *one* bank and find out the number of tellers it employs.

The point of this example is that whether or not a census study is *necessary* depends on the extent of variability among the population units with respect to the variable(s) about which data are needed in a research project.

In summary, two conditions must be met for a census study to be appropriate: (1) the feasibility condition and (2) the necessity condition. A census study will be *feasible* whenever a population is relatively small. A census study will be *necessary* only when the population units are extremely varied, that is, when each population unit is likely to be very different from all the other units. Consequently, the use of census studies in marketing is usually limited to industrial or institutional customer populations—that is, situations in which marketers may be interested in studying relatively small but highly varied populations. The next example illustrates the point.

Example. The XYZ Company is a major producer of a special kind of chemical that is used by 100 manufacturers of plastic products. Each of these 100 manufacturers accounts for a significant portion of XYZ's sales; moreover, each of them uses the chemical for totally different applications. Currently, XYZ is considering adding a new ingredient to its chemical that is supposed to improve the chemical's performance characteristics. But it first wants to get the reactions of its customers to the addition of the new ingredient. The situation faced by XYZ is appropriate for conducting a census study of its customers: The population is relatively small, and, more important, it consists of a very diverse group of manufacturers whose individual reactions are likely to be quite important to XYZ.

PROBABILITY VERSUS NONPROBABILITY SAMPLING

The subject matter of the rest of this chapter and the next two chapters relates to various methods for picking a sample of units from a given population. So let us preview two basic sampling methods that are available to researchers: (1) probability sampling and (2) nonprobability sampling.

A researcher using a probability-sampling method does not play a role in determining which *specific* population units are chosen to be part of a sample; that is, the researcher has no say in deciding whether or not each individual population unit is included in the sample. In probability sampling the researcher merely specifies some *objective scheme* for choosing units from a population. Once the scheme has been laid out, the selection of the sample units is independent of the personal preferences or biases of the researcher. Each population unit is assigned an objective probability of being selected on the basis of the specified sampling scheme (the probability of selection may or may not be the same for all population units, as we will see in the next chapter).

An intuitive explanation of sampling in general and probability sampling in particular is given in Research in Use 12.1, which contains excerpts from a brochure prepared by the Opinion Research Corporation, a well-known marketing research firm.

Definition. ***Probability sampling*** is an objective procedure in which the probability of selection is known in advance for each population unit.

Nonprobability sampling is more judgmental than objective in the sense that the selection of individual population units is not done on a strictly chance basis; a researcher's subjective judgment does play a role in determining which specific population units get included in the sample. Hence determining an objective probability of selection for each population unit beforehand is impossible.

Definition. ***Nonprobability sampling*** is a subjective procedure in which the probability of selection for each population unit is unknown beforehand.

A researcher conducting a sampling study has a choice from among a variety of probability and nonprobability methods. Exhibit 12.1 presents a classification of the basic types of probability and nonprobability sampling. These types are discussed in this and the next two chapters. The remainder of this chapter provides a detailed discussion of

Research in Use 12.1 *Opinion Research Corporation*

The Nature of Probability Sampling

THE KEY IDEA IN SAMPLING

You can draw accurate conclusions about the big world of reality from the little world of a sample.

Whether or not they know it, people use this idea in one way or another in everyday life.

The housewife knows how much salt to put in the soup by stirring the pot and tasting one spoonful.

Because she expects little difference between one sample and the next, she has confidence in her judgment as to the overall quality of the entire pot of soup, even though it is based on just one taste.

Our housewife also knows that strawberries at the top of the box may not be like those at the bottom. As a result, she is likely to examine a few berries from different parts of the box before deciding on the quality of the contents.

In attitude and market research the big world is the extraordinarily diverse and complex world of people: rural and urban, highly educated and poorly educated, rich and poor, skilled and unskilled, men and women, employers and employees, readers and non-readers, consumers and non-consumers. The list is endless.

The problem confronting the statistician is to construct a little world that is sufficiently like the big world to meet the objectives of the study. While there are other sampling techniques, probability sampling is almost always the most precise and efficient method of doing this.

RANDOM SELECTION

Probability sampling is based on random methods of selection. To most people, the term random suggests something haphazard and slipshod.

To the statistician, however, random selection is a precise mathematical concept that is applied according to a strict set of rules designed to give each person in a big world, or *universe* in the language of the statistician, a known probability of being selected in a sample. Thus in a study that employs probability sampling techniques, each person in the universe involved has a known chance of being represented in the survey results.

To rule out selective factors such as the interviewer's convenience, which might otherwise tend to bring too many easy-to-reach people into the sample, some random procedure is used to predesignate those specific individuals whom the interviewers are to contact.

(Continued)

Research in Use 12.1 *Continued*

AN EXAMPLE

A simple way to visualize the key idea in probability sampling is to imagine an army regiment lined up for review.

By selecting every fourth man listed on an alphabetical roster, we would have a probability sample in which each man in the regiment, or big world, had a known, in this case equal, opportunity to be selected—short soldiers as well as tall, those eager to be interviewed as well as those who are reluctant, married and unmarried soldiers, men who like their sergeants and men who don't. Moreover, the choice of who is to be interviewed would be independent of the subject of the interview.

If we interviewed the men selected, we then could use our sample, within mathematically definable limits, to draw conclusions about the whole regiment.

The civilian population, of course, cannot be lined up and sampled in this way. But by carefully following a detailed sampling plan, drawn as precisely as an engineer's blueprint, we have a practical and accurate method for sampling diverse and less easily manageable population groups.

Source: "Probability Sampling: A Simplified Explanation," a brochure published by the Opinion Research Corporation, 419 North Harrison Street, Princeton, NJ 08540-3574. Used by permission.

simple random sampling. The remaining, more advanced, probability-sampling methods are covered in Chapter 13. Nonprobability-sampling methods are discussed in Chapter 14. Factors playing a role in deciding whether to use probability or nonprobability sampling are also discussed in Chapter 14.

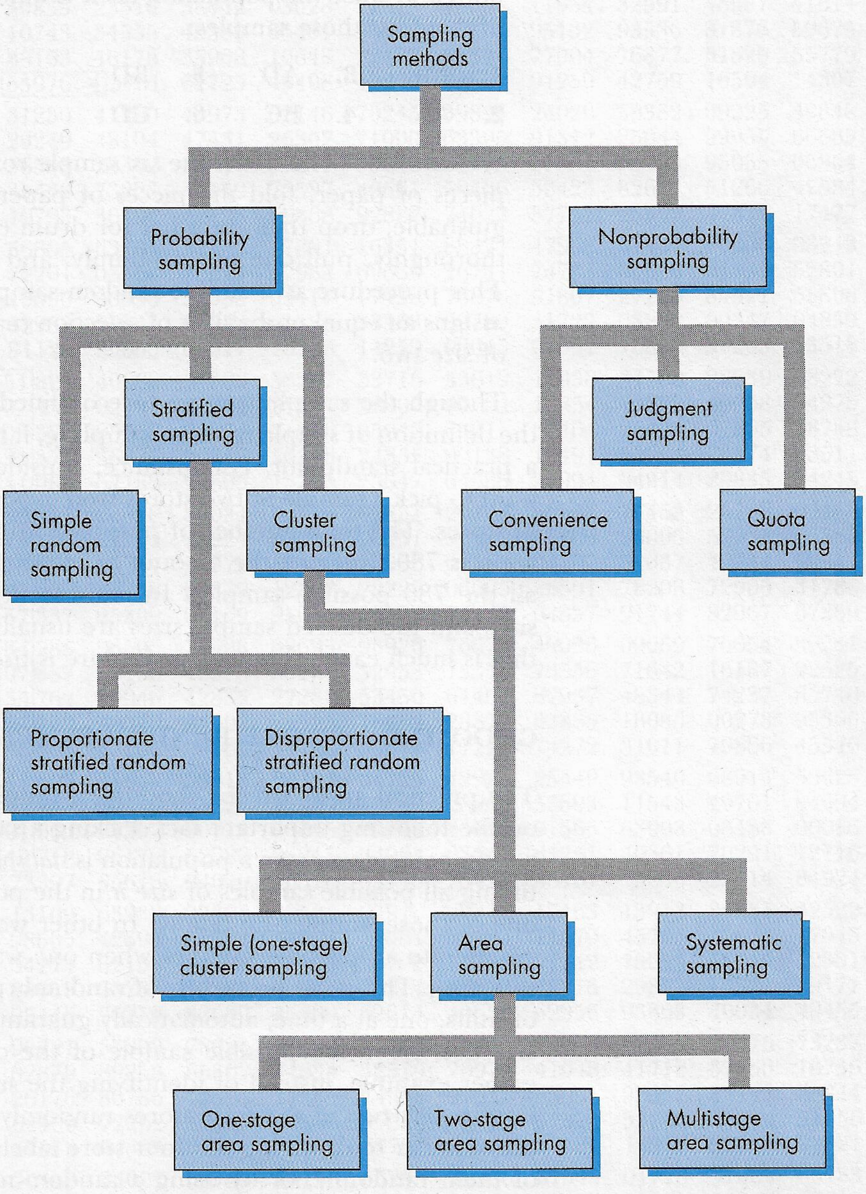
DEFINITION OF SIMPLE RANDOM SAMPLING

Simple random sampling is the most basic type of probability sampling for selecting a specified number of units from a population. A formal definition of simple random sampling is as follows:

Definition. ***Simple random sampling*** is a procedure in which every possible sample of a certain size within a population has a known and *equal* probability of being chosen as the study sample.

Example. Consider a hypothetical population containing just four stores: A, B, C, and D. Suppose we want to pick a sample of two stores from this population by using a simple-random-sampling procedure.

Exhibit 12.1 *Classification of Sampling Methods*


Sampling methods
Probability sampling
Nonprobability sampling
Stratified sampling
Judgment sampling
Simple random sampling
Cluster sampling
Convenience sampling
Quota sampling
Proportionate stratified random sampling
Disproportionate stratified random sampling
Simple (one-stage) cluster sampling
Area sampling
Systematic sampling
One-stage area sampling
Two-stage area sampling
Multistage area sampling

Let us first enumerate every possible sample of size two. Six different samples, each containing two stores, are possible within our population; and since the population itself is quite small, we can easily identify and list those samples:

1. AB **3.** AD **5.** BD
2. AC **4.** BC **6.** CD

We can now write down the six sample combinations on six identical pieces of paper, fold the pieces of paper so that they are indistinguishable, drop them in a hat (or drum or whatever), mix them up thoroughly, pull one out randomly, and use the sample indicated. This procedure is a simple-random-sampling procedure because it assigns an equal probability of selection (namely, ⅙) for every sample of size two.

Though the sampling procedure outlined in the example illustrates the definition of simple random sampling, it has serious drawbacks from a practical standpoint. For instance, consider a situation in which we want to pick a sample of two stores from a population of 40 (rather than 4) stores. The total number of possible samples of size two in this situation is 780. Imagine the amount of effort required to just enumerate all the 780 possible samples. In real-life situations, where population sizes and the desired sample sizes are usually large, a shortcut method that is much easier than this procedure is used, as discussed next.

CHOOSING A SIMPLE RANDOM SAMPLE IN PRACTICE

The practical approach for choosing a simple random sample is based on the following important fact: Picking a sample of size n by selecting n units at random from a population is *statistically equivalent* to first identifying all possible samples of size n in the population and then picking one of those samples randomly. In other words, one does not need to enumerate all possible samples when one wants to use simple random sampling. The easier procedure of randomly picking the desired number of units, one at a time, automatically guarantees an equal probability of selection for every possible sample of the desired size. Hence in our earlier example, instead of identifying the six possible samples, we can simply pick two of the four stores randomly—either by using a procedure similar to dropping the four store labels in a hat and drawing two of them randomly, or by using a random-number table, which is discussed next.

A random-number table consists of a group of digits that are arranged in random order; that is, any row, column, or diagonal in such a table contains digits that are sequenced in no systematic order. Table 12.1 is a table of random numbers. In this table the grouping of digits into 5-by-5 clusters is merely to facilitate reading and has no other significance.

Table 12.1 *Table of Random Numbers*

40743	39671	07812	42293	41539	40998	95829	11436	40890	16406
80833	23016	38342	03628	21313	46364	11934	82901	46587	41614
10743	34555	49339	36401	90000	77565	95132	93536	81876	59673
88103	46176	35968	10648	22427	22321	77904	76877	51829	55779
53976	43761	61725	49408	42200	98000	91259	42759	10504	14307
31230	41200	43973	62146	70245	39856	24029	38382	09225	49848
29249	48104	47431	26302	71090	63395	91312	25044	99955	66663
56588	28502	31317	37819	98119	50860	51639	55193	95055	06864
35438	73223	44149	70587	40561	50406	53423	82627	51266	42584
19793	30194	42198	59868	92882	74022	37797	93500	41877	17477
56865	63701	22341	37891	16351	88371	47968	97437	70400	93943
72201	03668	70611	47954	69660	26741	74733	00810	01152	52801
15890	96036	21583	04417	17438	76876	51807	27263	84521	55808
19983	27843	69758	61186	38368	39769	31722	68252	09347	04839
81126	95633	42617	16726	14259	08045	01691	16800	30293	23518
51823	40608	51380	58207	33716	53619	11038	31730	22239	08272
07454	92103	51564	21146	28807	00764	13854	73228	49968	94276
03375	95032	65271	71876	25947	84015	67464	28424	07995	48746
91374	36275	90717	56396	24855	99138	12491	52901	73174	48511
41086	67186	22605	15775	73647	64586	35994	54814	29835	64245
60431	24886	80034	96476	82831	20348	30157	73453	26122	09458
15899	81159	28224	17790	96251	30142	57206	30095	52935	43568
10052	05764	51779	47543	26866	02152	04707	54037	49744	59947
95657	52708	94879	12367	22228	06005	86391	78208	72906	31788
57337	25309	03169	95668	29124	14437	04837	91244	82967	67869
64988	98506	70398	68695	95970	76676	49086	09069	70654	39241
07387	98150	49313	08281	32958	15356	73356	71642	16187	92525
55762	30946	42835	27282	53450	01402	82937	48344	74237	85740
67720	48010	98907	44585	16464	28870	23885	18085	00278	93300
86803	62206	09160	81182	99671	17224	74272	31011	79866	45546
41693	81849	08818	65391	64726	42923	25449	98540	68914	56626
82145	50140	45227	37427	70649	26962	55693	11543	49701	84694
01600	81297	00713	88645	02774	44711	91565	63908	06488	00086
42718	96176	52680	93215	04964	03641	61251	40361	20291	12116
73397	35616	04020	70326	93138	26497	86161	13730	22414	84924
13100	63938	07302	89521	07910	03230	17163	13913	83881	42522
98505	42693	34714	41842	12051	05483	11279	45705	10644	17043
04274	62158	83986	36968	05792	14755	91042	78940	58215	92301
68619	63767	65875	39595	78214	45903	52216	29857	14521	79177
95445	76259	32639	45690	62614	40656	62956	93663	40034	49455
96721	33899	78824	46185	36809	45441	86203	37607	91375	73237
02879	89258	65507	31638	32510	76979	41048	41712	37650	16736
26176	86786	12324	17380	24938	97032	45054	98632	05148	78244
62199	72262	44802	58198	95923	36412	67574	33290	25200	51556
47013	04898	63931	28253	62793	81973	20581	18977	98348	93521
30448	34027	58753	78128	05239	72236	67343	01778	97089	63183
10561	14780	82473	64810	81448	37989	15817	81860	32942	57614
99651	20030	31874	43927	05097	94280	01378	48137	13308	43065
32483	25721	47734	00809	00371	37366	11124	70027	46199	44451
26341	11426	78027	49249	00137	92846	77094	38012	42631	93977

Source: Reprinted from Page 349 of A MILLION RANDOM DIGITS WITH 100,000 NORMAL DEVIATES, by The Rand Corporation, New York: The Free Press, 1955. Copyright 1955, 1983 by The Rand Corporation. Used by permission.

Let us see how we can use this table to select two stores from our population of four stores. Our selection procedure consists of the following steps:

1. Assign consecutive numbers to the stores in the population list (i.e., sampling frame) as follows:

1	**A**
2	**B**
3	**C**
4	**D**

2. Randomly start with some digit in the table.
3. Read down (or up) the column (or row) containing the digit.
4. Pick the stores corresponding to the digits that are read until two stores are selected.

For instance, suppose we randomly start (e.g., by stabbing with pencil point at the table without looking at it) with the fifth row in the first column of the table and decide to read across the row. The starting digit is 5. There is no number as large as 5 in our population list since there are only four stores in it. We simply move on to the next digit on the row which is 3, and corresponds to store C. We pick C and move on to the next digit, which is 9; we ignore this digit since it does not correspond to any store in our population. We proceed in a similar fashion until we have a sample size of two. The random numbers selected for this particular sequence are 3 and 4. The sample derived from this sequence consists of stores C and D. The sequence of random numbers that yields this sample is boxed in the table.

As another illustration of using the random-number table, consider a situation in which we want to pick a simple random sample of 10 counties from a population containing 60 counties. Here we number the counties 00 through 59 (the starting number 00 is chosen arbitrarily; the first population element can very well be designated as 01 or 10—or any other number, for that matter). We randomly start with any two-digit number in the table, say, the first two digits in row 16 of the table, which are 51. If we now read down this two-digit column, our simple random sample of 10 counties contains those numbered 51, 07, 03, 41, 15, 10, 57, 55, 01, and 42. The sequence of random numbers that yields this sample is also boxed in the table.

In this computer age a researcher can obtain electronic help in choosing a simple random sample. Many computers have pseudo-random-number generators that can readily provide a set of random numbers within any desired range. Unless the sample size is relatively small, computerized selection of a simple random sample is faster and easier than manual selection using a random-number table. Computer selection is particularly helpful in telephone surveys requiring a random sample of respondents. Nowadays, special computer programs are available to perform random-digit dialing, in which telephone numbers are randomly selected and dialed by a computer.

The preceding sample selection procedure assigns an equal probability of selection for *every unit* in the population. Indeed, a key property of simple random sampling is that every population unit is just as likely to end up in the chosen sample as any other unit. So why is simple random sampling not defined as "a procedure in which each population unit has a known and equal chance of being chosen for the sample"? This definition is somewhat simpler than the formal definition given earlier. However, the reason for not defining simple random sampling in this way is that the property of each population unit's having an equal probability of selection is not unique to simple random sampling. There are other probability-sampling methods (e.g., proportionate stratified random sampling, to be discussed in the next chapter) that also possess this property. Hence the apparently simpler definition does not unambiguously portray the key statistical process underlying simple random sampling—that is, the process of guaranteeing the same probability of selection for *every sample of a given size,* which is unique to simple random sampling.

SAMPLING ERROR AND SAMPLING DISTRIBUTION

The basic purpose of sampling is to generate sample data that can be used to infer something about the nature of the population. For instance, from data collected from a cross section of families living in a city, we can estimate the mean income of the population of families in the city, the proportion of families with microwave ovens, and the like. However, unless we conduct a census study (i.e., a study of *all* the families in the city), we cannot pinpoint the actual mean family income, the exact proportion of families with microwave ovens, and so on. The reason we cannot do so is the presence of sampling error, which is discussed next.

Sampling Error

The actual, or true, population mean value or population proportion for any variable (e.g., income, product ownership) is referred to as a ***parameter.*** An estimate of it from sample data is referred to as a ***statistic.*** Whenever a sampling procedure is used to estimate a population parameter, there may be some discrepancy between the sample statistic and the actual (but unknown) parameter value, no matter how objectively and carefully the sample is chosen. Such a discrepancy is commonly called sampling error.

Parameter *A summary measure based on data from all units in a given population.*

Statistic *A summary measure based on data from a sample of units.*

Definition. ***Sampling error*** is the difference between a statistic value that is generated through a sampling procedure and the parameter value, which can only be determined through a census study.

Table 12.2 *Expenditures for Eating Out for a Hypothetical Population*

Family Number	Annual Expenditure for Eating Out ($)
1	50
2	100
3	150
4	200
5	250
6	300
7	350
8	400
9	450
10	500

Knowledge of the *magnitude* of the sampling error is essential for ascertaining how precisely the population parameter can be estimated from a sample statistic value. But we cannot accurately determine the magnitude of the sampling error associated with a sampling process for two reasons: (1) We do not know the true population parameter value (if we already did, there would be no need for any sampling study), and (2) the sample statistic value itself may vary from sample to sample within the same population (more will be said on this variation in the next section). Hence we can only estimate the average amount of sampling error associated with a given sampling procedure. An important statistical concept that is critical for estimating sampling error is sampling distribution, which is discussed next.

Sampling Distribution

Consider a population consisting of just 10 families. Suppose the annual expenditures of the 10 families for eating out are as shown in Table 12.2. Here the population mean expenditure for eating out (i.e., the population *parameter* value) is $275. Suppose we want to estimate the population parameter on the basis of data from a simple random sample of just 2 families.

There are 45 possible samples of size two in this population. For any sample the value of the sample mean expenditure for eating out (i.e., the sample *statistic* value) will depend on the families included in it. Table 12.3 contains a partial list of the various samples of size two, along with their sample mean expenditures for eating out. Notice that the sample mean values span a range—from $75 to $475—and that some mean values (such as $275) occur more frequently than others, since a greater number of samples yield those mean values. Thus, theoretically, if we

Table 12.3 *Partial List of Possible Samples and Sample Means*

Samples of Two Families	Sample Mean Values ($)
1, 2	75
1, 6; 2, 5; 3, 4	175
1, 10; 2, 9; 3, 8; 4, 7; 5, 6	275
5, 10; 6, 9; 7, 8	375
9, 10	475

were to pick samples of size two over and over again from this population and compute the sample mean value each time, we would observe a range of mean values, with certain values showing up more often than others. This phenomenon underlies the sampling distribution concept.

Definition. A ***sampling distribution*** is a representation of the sample statistic values—obtained from every conceivable sample of a certain size chosen from a population by using a specified sampling procedure—along with the relative frequency of occurrence of those statistic values.

Exhibit 12.2 represents the sampling distribution associated with a simple-random-sampling procedure for picking samples of 2 families from our population of 10 families. The horizontal axis in this exhibit portrays the actual range of sample mean values, and each vertical line indicates the relative frequency of occurrence of a corresponding sample mean value. For instance, if we pick many samples of size two and compute the sample mean for each, approximately $3/45$ (or about 6.7%) of those values will be $175. That is, three independent samples yield a sample mean value of $175 (see Table 12.3). And since each of these samples has the same selection probability (namely, $1/45$), the relative frequency (or probability) of occurrence for a sample mean value of $175 is given by

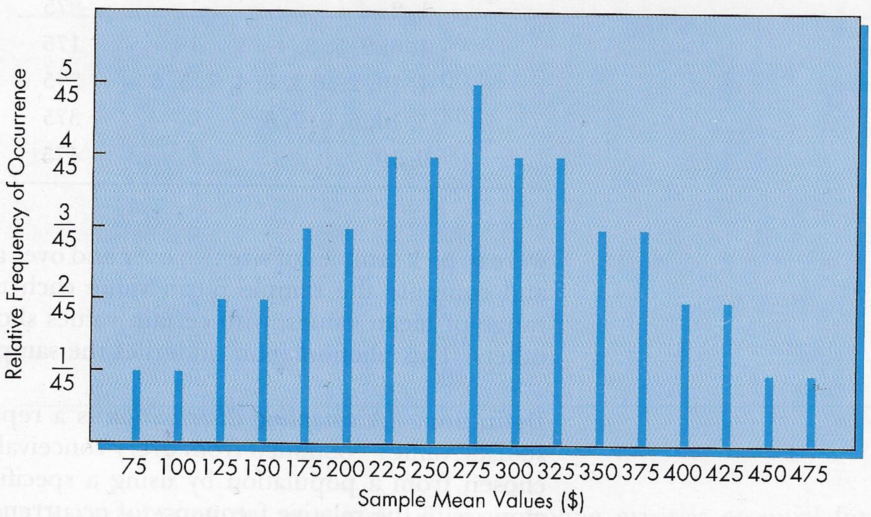
$$\tfrac{1}{45} + \tfrac{1}{45} + \tfrac{1}{45} = \tfrac{3}{45}.$$

We emphasize that the sampling distribution concept is a theoretical concept in the sense that in real life we will only pick *one* random sample, rather than a large number of random samples that are needed to construct a sampling distribution. Nevertheless, an understanding of this concept is essential since it forms the basis for estimating a population parameter from a sample statistic value. The nature of the sampling distribution associated with a sampling procedure is the key to estimating the sampling error generated by the procedure.

The sampling distribution diagramed in Exhibit 12.2 has two noticeable features:

1. The most frequently occurring sample mean values cluster around the population parameter value of $275.

Exhibit 12.2 *Sampling Distribution for Simple Random Samples of Two Units*



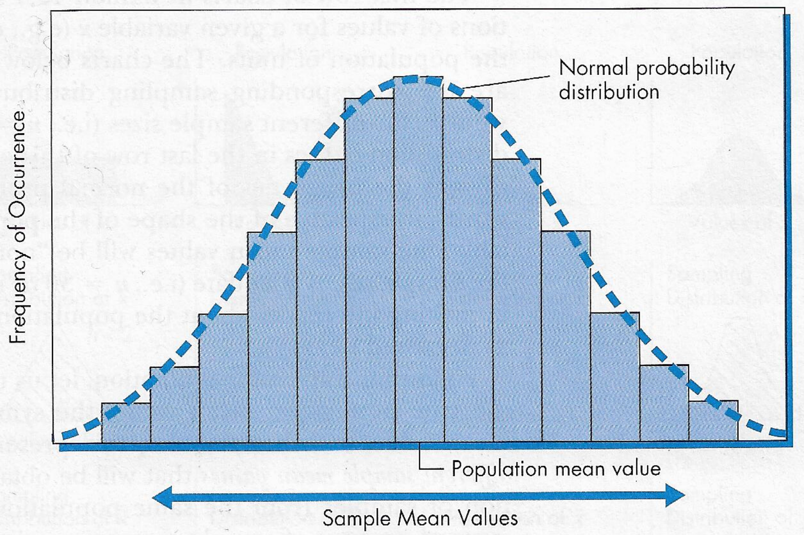
2. Extreme sample mean values, although they are possible, are not very likely to occur.

These features are common to all sampling distributions. Repeated selection of random samples from the same population will yield sample statistic values such that the closer a value is to the population parameter value, the more often it will occur.

In our hypothetical illustration the population size of 10 families is relatively small and hence restricts the total number of possible sample mean values. For this reason the sampling distribution shown in Exhibit 12.2 is discrete and has a somewhat spiked appearance. If the population size is larger and more varied (as is normally the case in real life), a greater variety of sample compositions, and hence a greater variety of sample mean values, will be possible. Consequently, the sampling distribution will contain a much larger number of vertical lines and will have a denser appearance than the one shown in Exhibit 12.2.

When an infinite number of sample compositions are possible, the range of sample mean values can be considered to be *continuous* rather than *discrete;* that is, the sample mean can be considered as being able to take on any value in the range rather than only certain specific values, like $75, $100, and $125 in our example. The sampling distribution for a continuous variable is usually shown as a *histogram* rather than as a diagram like Exhibit 12.2 that indicates the probability of occurrence of *every possible* value of the variable. An example of a histogram repre-

Exhibit 12.3 *Sampling Distribution Shown as a Histogram*



senting the sampling distribution of the sample mean is shown in Exhibit 12.3.

In Exhibit 12.3 the height of each rectangle denotes the total frequency of occurrence of all sample mean values included within the range defined by the two vertical sides of the rectantle. Thus the histogram portrays the frequency of occurrence for different *ranges* of sample mean values rather than for *each possible* sample mean value. However, notice that the probability distribution diagram in Exhibit 12.2 and the histogram in Exhibit 12.3 have the same general shape. Indeed, the histogram can be approximated by a bell-shaped curve, as shown by the dotted line in Exhibit 12.3. Such a smooth curve is sometimes called the *theoretical* sampling distribution.

Central Limit Theorem
A theorem stating that for a sufficiently large sample size (in practice, a sample size of 30 or more), the sampling distribution curve for sample means associated with a sampling procedure will be centered on the population parameter value and will have all the properties of a normal probability distribution.

The sampling distribution curve has an important characteristic based on a theorem called the ***central limit theorem***: For a sufficiently large sample size (in practice, a sample size of 30 or more), the sampling distribution curve for sample means associated with a sampling procedure will be centered on the population parameter value and will have all the properties of a *normal probability distribution* (i.e., the so-called normal curve).[4] This characteristic holds *regardless* of how the values of the original variable are distributed in the population, as Exhibit 12.4 demonstrates.

---

[4] For a discussion of the properties of a normal probability distribution, see any introductory statistics textbook, such as D. R. Anderson, D. J. Sweeney, and T. A. Williams, *Introduction to Statistics—An Applications Approach* (St. Paul, Minn.: West, 1981).

The first row of charts in Exhibit 12.4 shows four different distributions of values for a given variable x (e.g., expenditure on eating out) in the population of units. The charts below each population distribution are the corresponding sampling distributions of the variable's mean value $\bar{x}$, for different sample sizes (i.e., $n = 2$, 5, and 30). The sampling distribution curves in the last row of this exhibit are all bell-shaped and possess the properties of the normal probability distribution. In other words, irrespective of the shape of the population distribution of a variable, the sample mean values will be "normally" distributed as long as the sample size is adequate (i.e., $n = 30$ or more). This property is useful in making inferences about the population on the basis of sample data, as discussed next.

Following conventional notation, let us use the symbol μ to represent the true population mean value, the symbol $\bar{x}$ to represent a sample mean value, and the symbol $\sigma_{\bar{x}}$ to represent the *standard deviation of the different sample mean values* that will be obtained through repeated selection of samples from the same population. We can interpret $\sigma_{\bar{x}}$ as an average amount of sampling error associated with the sampling procedure. Traditionally, it is referred to as the ***standard error*** of the mean.

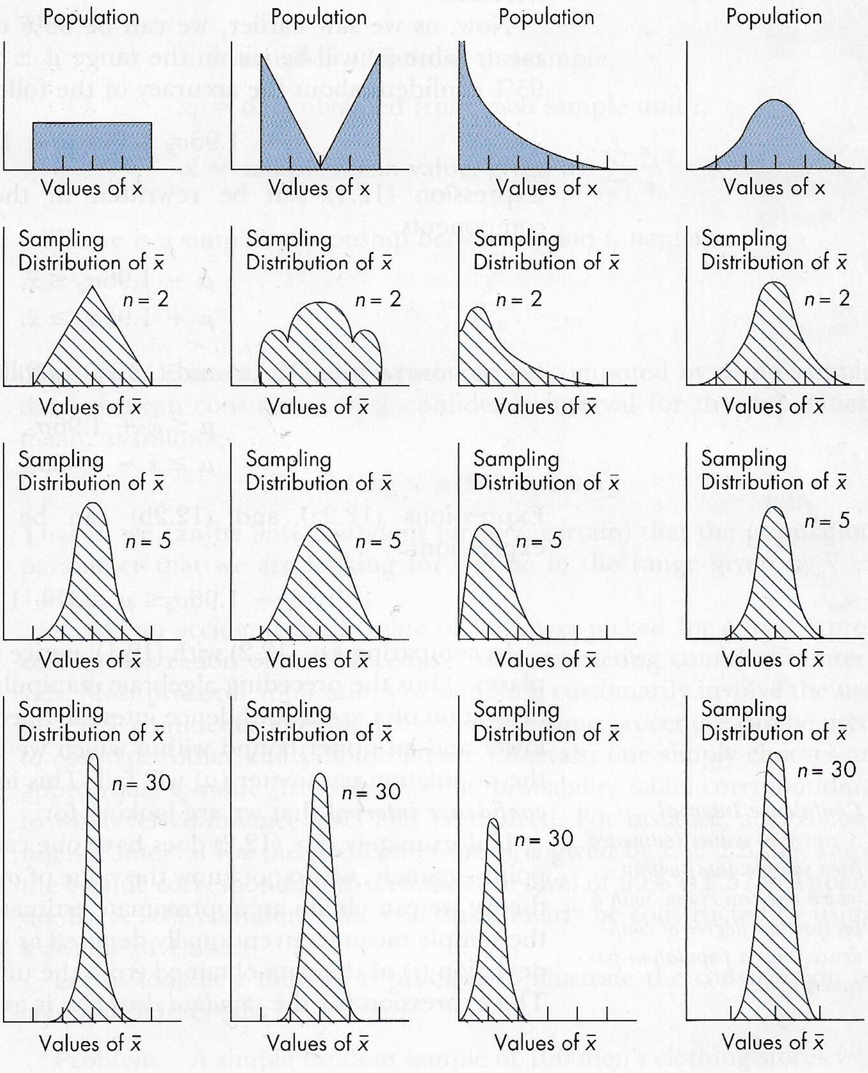
Standard Error
An indicator of the extent of sampling error; it represents the standard deviation of sample statistic values that will be obtained through repeated sampling from a population by using the same procedure.

If the theoretical sampling distribution behaves like the normal probability distribution, a *confidence statement* like the following can be made: "If we repeatedly pick random samples of a given size from a population, we can predict that a certain percentage (say, q%) of the sample mean values will be within the range given by $\mu \pm z_q\sigma_{\bar{x}}$," where z_q is the standard normal-deviate value corresponding to q%. The z-values can be read from a normal probability table such as the one shown in Appendix B. Alternatively, the following statement can also be made: "If we pick *one* random sample of a given size from a population, we can predict with q% confidence (or certainty) that the mean value for that sample will be within the range given by $\mu \pm z_q\sigma_{\bar{x}}$." This latter statement forms the basis for estimating confidence intervals for population parameter values.

CONFIDENCE-INTERVAL ESTIMATION

The preceding section illustrated that certain confidence statements can be made about sample statistic values given that the population parameter value (μ) and the standard error ($\sigma_{\bar{x}}$) are known. However, in practice, neither μ nor $\sigma_{\bar{x}}$ is known. Moreover, our purpose is to predict, with a certain degree of confidence, the range of values within which the population parameter is likely to be, given data from just *one* random sample. We can accomplish this purpose by drawing upon the theoretical principles discussed in the preceding sections and by making some approximations.

Exhibit 12.4 *Sampling Distributions for Different Populations and Sample Sizes*



Source: Ernest Kurnow, Gerald J. Glasser, and Frederick R. Ottman, *Statistics for Business Decisions* (Homewood, Ill.: Richard D. Irwin, 1959), pp. 182-183. Used by permission.

Confidence Intervals for Population Mean Values

The *z*-value corresponding to a *two-tailed* *q*-value of 95% is 1.96 (see Appendix B). Notice the emphasis on *two-tailed;* this term underscores that a confidence interval is symmetric about the mean (or proportion), and therefore the z-value corresponds to the tail portion of a normal curve covering an area equivalent to $(100 - q) \div 2$. This two-tailed nature of confidence intervals is *implied* throughout the discussion hereafter.

Now, as we saw earlier, we can be 95% confident that any sample mean value ($\bar{x}$) will be within the range $\mu \pm 1.96\sigma_{\bar{x}}$. That is, we can be 95% confident about the accuracy of the following expression:

$$\mu - 1.96\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1.96\sigma_{\bar{x}}. \tag{12.1}$$

Expression (12.1) can be rewritten in the form of two separate components:

$$\mu - 1.96\sigma_{\bar{x}} \leq \bar{x}, \tag{12.1a}$$

$$\mu + 1.96\sigma_{\bar{x}} \geq \bar{x}. \tag{12.1b}$$

Let us now rewrite (12.1a) and (12.1b) as follows:

$$\mu \leq \bar{x} + 1.96\sigma_{\bar{x}}, \tag{12.2a}$$

$$\mu \geq \bar{x} - 1.96\sigma_{\bar{x}}. \tag{12.2b}$$

Expressions (12.2a) and (12.2b) can be combined into a single expression:

$$\bar{x} - 1.96\sigma_{\bar{x}} \leq \mu \leq \bar{x} + 1.96\sigma_{\bar{x}}. \tag{12.2}$$

In comparing Eq. (12.2) with (12.1), notice that $\bar{x}$ and μ have switched places. Thus the preceding algebraic manipulations have resulted in the derivation of a *useful* confidence interval. Specifically, we have derived a lower and an upper bound within which we can be 95% confident that the population parameter (μ) will fall. This interval is really the type of ***confidence interval*** that we are looking for.

Confidence Interval
A range of values (estimated from sample data) within which one can claim, with a prespecified degree of confidence, that a population parameter falls.

Unfortunately, Eq. (12.2) does have one catch from a practical standpoint—namely, we do not know the value of $\sigma_{\bar{x}}$. However, from statistical theory we can obtain an approximate estimate of the standard error of the sample mean (conventionally denoted as $s_{\bar{x}}$) if we know the standard deviation (s) of the data obtained from the units in our random sample. The expression for the *standard deviation* is as follows.[5]

---

[5] A few remarks about the expression for the standard deviation are in order. First, the $n - 1$ in the denominator (rather than n) ensures that the sample standard deviation is an *unbiased* estimate of the population standard deviation. Second, a "finite population correction factor" is applied to this expression when the sample size is relatively large compared with the population size. Specifically, when the sample size (n) is more than about 10% of the population size (N), the expression for s is multiplied by the factor $\sqrt{(N - n)/(N - 1)}$, resulting in a lower value for s. Third, more convenient computation formulas are available if one wants to calculate the sample standard deviation by hand. Further information on these issues can be found in most basic statistics textbooks; see, for instance, Anderson, Sweeney, and Williams, *Introduction to Statistics*.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}},$$

where

n = number of units in the sample,

x_i = data obtained from each sample unit i,

$\bar{x}$ = sample mean value, given by $\sum_{i=1}^{n} \frac{x_i}{n}$.

There is a simple relationship between $s_{\bar{x}}$ and s, namely,

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}.$$

Thus given $\bar{x}$ and $s_{\bar{x}}$ (both of which can be computed by using sample data), we can construct a 95% confidence interval for the population mean, as follows:

$$\bar{x} - 1.96 s_{\bar{x}} \leq \mu \leq \bar{x} + 1.96 s_{\bar{x}}.$$

That is, we can be 95% confident (or 95% certain) that the population parameter that we are looking for will be in the range given by $\bar{x} \pm 1.96 s_{\bar{x}}$.

It was no accident that a value of 95% was picked for q in the preceding illustration of the procedure for constructing confidence intervals. Most practical applications of sampling customarily involve the use of a 95% confidence interval. However, the same procedure can be used to construct other kinds of confidence intervals; one simply chooses an appropriate z_q-value (from the normal probability table) corresponding to whatever confidence level may be desired. For instance, a 99% confidence interval for the population mean is given by $\bar{x} \pm 2.575 s_{\bar{x}}$, since the z-value corresponding to a confidence level of 99% is 2.575 (Appendix B). A 90% confidence interval can similarly be constructed by using a z-value of 1.645.

Let us look at a numerical problem to illustrate the construction of confidence intervals.

Problem. A simple random sample of 100 men's clothing stores was chosen from the population of such stores in a city. The average annual sales of men's suits in this sample was 1278 units, and the standard deviation of sales was 399 units. Construct a 95% confidence interval for the average annual sales of men's suits in the *population* of stores.

In this problem $n = 100$, $\bar{x} = 1278$ units, and $s = 399$ units. Thus

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{399}{\sqrt{100}} = \frac{399}{10} = 39.9 \text{ units}.$$

The 95% confidence interval is

$$\begin{aligned}\bar{x} \pm 1.96s_{\bar{x}} &= 1278 \pm (1.96)(39.9)\\ &= 1278 \pm 78.204\\ &= 1278 \pm 78, \text{ approximately.}\end{aligned}$$

From the sample data we can be 95% confident that the average annual sales of men's suits, across *all* men's clothing stores in the population, is between 1200 and 1356 units.

Confidence Intervals for Population Proportions

The parameters of interest in some research projects may include population *proportions* (e.g., proportion of households in a city that own video recorders, proportion of firms in a region that employ more than 50 people). In such instances one must construct confidence intervals for population proportions from sample data. The procedure for determining such confidence intervals is quite similar to the one we discussed for population mean values. However, the symbolic notation used and the computation of the standard error are somewhat different. The following notation is customary for confidence intervals for population proportions:

π = true population proportion (i.e., the parameter value),
p = proportion obtained from a single sample (i.e., the statistic value),
s_p = estimate of the standard error of the sample proportion.

The 95% confidence interval for the population proportion is given by

$$p - 1.96s_p \le \pi \le p + 1.96s_p, \tag{12.3}$$

where

$$p = \frac{\text{number of sample units having a certain feature}}{\text{total number of sample units (i.e., } n)}$$

and

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

Notice that we only need to know the sample proportion and the sample size to be able to estimate the standard error s_p.

Problem. A simple random sample of 100 grocery stores was chosen from *all* grocery stores in a city. Only 64 of the 100 stores carried potted plants; that is, the proportion of stores in the sample that carried potted plants was .64. Construct a 95% confidence interval for the *population* proportion of stores carrying potted plants.

In this problem $n = 100$ and $p = .64$. So

$$s_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.64)(.36)}{100}} = .048.$$

The 95% confidence interval is, then,

$$\begin{aligned} p \pm 1.96s_p &= 64 \pm (1.96)(.048) \\ &= .64 \pm .09408 \\ &= .64 \pm .09, \text{ approximately.} \end{aligned}$$

This confidence interval can also be expressed in *percentage* terms: 64% ± 9%. In other words, we can be 95% confident that between 55% and 73% of all grocery stores in the city carry potted plants.

DETERMINING SAMPLE SIZE

A basic purpose of sampling is to estimate population parameter values as accurately as possible. We have already seen that we cannot know for sure what a population parameter value is unless we do a census study. The only way to "predict" a parameter value from sample data is to come up with an estimate in the form of a confidence interval, usually the 95% confidence interval. For a given confidence level the tighter, or more compact, the confidence interval, the more *precise*, and hence more useful, the interval estimate will be.

For instance, consider two sampling studies, say A and B, that are conducted to estimate the proportion of a target market that has tried a recently introduced product. Suppose the results of the two studies are as follows: From study A one can be 95% confident that the proportion of all target customers who have tried the product is between .48 and .52; from study B one can be 95% confident that the proportion of all target customers who have tried the product is between .35 and .65. Notice that the two confidence intervals imply that the sample proportion obtained through each study is the same, namely, .50.

If you are a product manager, which of these two studies will you consider to be more useful for decision-making purposes? Study A is apparently superior, because the interval estimate derived from study A is more precise. It implies that you can be reasonably certain that about half your target market has tried your new product, thereby providing you with a precise evaluation of the effectiveness of your marketing strategy for your product. The interval estimate obtained from study B contains little, if any, useful information. It merely indicates that you can be reasonably certain that the proportion of your target market that has tried your new product can be as low as .35 or as high as .65. This information is not really helpful in evaluating the effectiveness of the marketing strategy to launch the product.

In short, to be managerially useful, a confidence-interval estimate must be as compact as possible. That is, the term that is subtracted from and added to the sample statistic value (namely, the ± term) must be as small as possible. This point has key implications regarding certain factors that play a role in sample size determination, as discussed next.

Factors Influencing Sample Size

Let us examine the composition of the ± term that is used in the construction of confidence intervals. This term is $z_q s_{\bar{x}}$ when we are dealing with population mean values, and it is $z_q s_p$ when we are dealing with population proportions. We know that $s_{\bar{x}}$ = (sample standard deviation) $\div \sqrt{n}$. The expression for s_p is $\sqrt{p(1-p)} \div \sqrt{n}$; however, $\sqrt{p(1-p)}$ is nothing but the sample standard deviation for a proportion. Hence, irrespective of whether we are interested in constructing confidence intervals for population means or population proportions, the general expression for the ± term is

$$z_q \times \frac{\text{sample standard deviation}}{\sqrt{n}},$$

where z_q is the standard normal deviate corresponding to the desired confidence level and n is the sample size.

This expression for the ± term is central in determining how compact or how wide a confidence interval will be. By simply examining this expression, we can gain some insights about factors that play a role in determining the sample size for a study.

***Precision Level** A measure of the degree of compactness of a confidence interval. The narrower the confidence interval, the higher the precision level is.*

First, the desired degree of compactness of the confidence interval—usually known as the desired ***precision level***—plays a role in determining sample size. Other things remaining the same, the greater the desired precision level, the smaller the acceptable magnitude for the ± term is, and hence the larger the sample size should be.

Given a certain desired precision level, the larger the values of z_q, the larger the sample size should be. That is, if the numerator of the expression is large, the denominator must be correspondingly large in order to keep the magnitude of the expression small. Thus a second factor that must be considered in determining sample size is the z_q-value, or the desired *confidence level*, on which the value of z_q depends. Other things remaining the same, the more confidence we want to have in the interval estimate, the larger the z_q-value will be, and hence the larger the sample size should be. For instance, the z-values associated with the 95% and 99% confidence levels are 1.96 and 2.575, respectively. Thus if we desire a 99% confidence interval (rather than a 95% confidence interval) but want it to have the *same* width, or degree of precision, as a 95% confidence interval, we need to pick a larger sample than what we would need for constructing the 95% confidence interval.

The sample standard deviation is another key variable in the expression for the ± term. What does a large or a small standard deviation mean? Standard deviation is a summary measure of how much the data provided by individual sample units differ from the overall central tendency of the total sample (i.e., the sample mean or proportion). The standard deviation of a sample is a measure of how diverse or similar the individual units that make up the sample are; the larger the standard deviation, the greater is the diversity of the units included in the sample. Since a simple random sample is supposed to be an objective representation of the population, we can view the sample standard deviation as being an approximate measure of the extent of variability or similarity of units in the population itself. Hence a third factor that must be considered in determining sample size is the *degree of variability* across the population units. Other things remaining the same, the greater the degree of variability within the population, the larger the sample size should be.

Last, but not least, the *resources available* (e.g., time and money) for conducting a study must also be considered in determining sample size. The available resources are quite critical because, in practice, they place an upper limit on sample size. After all, if resources are unlimited, we can simply pick the largest possible sample, or even do a census study, and not even worry about determining sample size. Such an ideal situation rarely, if ever, exists in real life. Although decision makers may desire a certain level of confidence in research results and want the results to have a specified degree of precision, the fulfillment of their desires often is constrained by a lack of resources. In many real-life situations the budget available becomes the sole determinant of sample size.

Sample Size Determination Methods

In the preceding section we discussed, in a qualitative fashion, the various factors that can influence sample size determination. This section focuses on using quantitative estimates of some of those factors to determine numerical values for sample sizes in different situations.

An important point we must discuss at the outset is that the estimation of numerical values for sample sizes requires a researcher to make certain assumptions and provide certain inputs that are crucial to the meaningfulness of the estimated sample sizes. A researcher, in collaboration with the relevant decision maker, must provide the following inputs in order to determine the sample size for a study:

1. The desired *precision level,* that is, the magnitude of the ± term that the researcher (or decision maker) is willing to tolerate. For example, a decision maker may want to estimate the mean age of a target market within four years, or the decision maker may want to estimate the percentage of a population that owns microwave ovens within a margin of error of ±5%.
2. The desired *confidence level,* that is, the degree of confidence that the decision maker wants to have in the interval estimate.

3. An *estimate* of the degree of variability in the poplation, expressed in the form of a *standard deviation.*

Given a desired precision level (customarily denoted as H to signify *half* interval—i.e., one-half of the confidence-interval range), a desired confidence level (q), and an estimate of the standard deviation (s), we can write the following equation:

$$H = \frac{z_q s}{\sqrt{n}}. \tag{12.4}$$

We can square both sides of Eq. (12.4) and rewrite it as follows:

$$n = \frac{z_q^2 s^2}{H^2}. \tag{12.5}$$

Equation (12.5) can be used to determine sample size, irrespective of whether the objective is to estimate a population mean or a population proportion. The following problem demonstrates its application in a situation involving estimation of a population mean.

Problem. A marketing manager of a frozen-foods firm wants to estimate the average annual amount that families in a certain city spend on frozen foods per year. He wants the estimate to be within ±\$10. When such an interval estimate is constructed, he wants to be able to have 99% confidence in it. He estimates that the standard deviation of annual family expenditures on frozen foods is about \$100. How many families must be chosen for this study?

In this problem $H = \$10$, $s = \$100$, and $z_q = 2.575$ (corresponding to a confidence level of 99%). Notice that H and s are expressed in the *same units,* namely dollars; H and s must be specified in the same units before solving any problem of this kind. We can solve for the value of n by using Eq. (12.5):

$$n = \frac{(2.575)^2(100)^2}{(10)^2} = 663 \text{ families, approximately.}$$

Although the preceding numerical solution to our problem is straightforward, there is a critical catch hidden in the problem; this catch relates to the estimate of the standard deviation of expenditures on frozen foods. How did the manager obtain this estimate? If it is merely an educated guess, how much trust can we place on its accuracy? What impact will this guessed value have on the sample size estimate? The impact can be great. For instance, even if the manager's estimate falls short of the true standard deviation by a mere \$10 (i.e., if the true value of the standard deviation is \$110 instead of the assumed value of \$100), the actual sample size required will be approximately 802 families, and hence the estimated sample size of 663 will be off by 139 families.

The point is that the estimated sample size is very sensitive to even minor errors in the assumed value of the standard deviation. In many real-life situations, where estimates of standard deviations are often merely based on intuitions, using Eq. (12.5) to estimate sample sizes is meaningless.

Can a manager or a researcher do better than estimate the standard deviation through subjective judgment? Better estimates are available under certain circumstances:

1. If a study similar to the one being contemplated has been done in the past, the researcher may be able to use the standard deviation obtained from that study.
2. If enough time and money are available, a pilot study using, say, 30 or so units can be conducted simply to estimate the standard deviation of the variable.[6]
3. If the minimum and maximum values of the variable in the population are known, *and* if the variable values for the population units can be assumed to be *normally distributed,* the standard deviation can be estimated as follows:

$$s = \frac{\text{maximum value} - \text{minimum value}}{6}.$$

This expression is derived as follows: Given a group of numbers that are normally distributed, almost all the numbers will be included in the interval formed by the group mean ± 3 standard deviations. In other words, the difference between the largest and smallest numbers in the group should be approximately equal to six standard deviations.

Though the three circumstances outlined here may yield better estimates of the standard deviation than those based on subjective judgment, a researcher would be wise to bear in mind that such estimates are still only approximations and cannot be expected to be totally error-free.

Let us now consider sample size determination in situations involving estimation of population proportions. As mentioned earlier, Eq. (12.5) can be used to determine sample size when the objective is to estimate a population proportion. However, estimating the standard deviation for this parameter is perhaps even more complex than estimating the standard deviation for a population mean. By definition, the standard deviation of a proportion is expressed in terms of the proportion itself; that is, $s = \sqrt{p(1-p)}$. Hence to estimate s, we need to estimate p, which is what the study, when completed, is supposed to find out.

---

[6] For a discussion of biases that may occur in using this procedure and how to correct for them, see Ronald E. Shiffler and Arthur J. Adams, "A Correction for Biasing Effects of Pilot Sample Size on Sample Size Determination," *Journal of Marketing Research,* 24 (August 1987), pp. 319–321.

Of course, if resources permit, one can conduct a pilot study to estimate p and then use that estimate to determine the sample size for the final study. However, a more meaningful approach to use in problems involving proportions is to determine an upper limit for the sample size that can guarantee an interval estimate that has the desired precision level *and* confidence level. How can we determine this upper limit? If we examine the general sample size formula given by Eq. (12.5), we see that for given values of H and z_q, n will take on its maximum value when s assumes its maximum value. For proportions the value of s is largest when $p = .5$, and that largest value is given by

$$\begin{aligned} s &= \sqrt{.5(1-.5)} \\ &= \sqrt{(.5)(.5)} \\ &= .5. \end{aligned}$$

Hence the upper limit for sample size, $n_{\max}$, can be computed by using the following equation:

$$n_{\max} = \frac{.25z_q^2}{H^2} \qquad (12.6)$$

Problem. A sporting goods marketer wants to estimate the proportion of tennis players among high school students in the United States. The marketer wants the estimate to be accurate within ±.02 and wants to have 95% confidence in the interval estimate. A pilot telephone survey of 50 high school students showed that 20 of them played tennis. Estimate the required sample size for the final study from the given data. What should the sample size be if the desired precision and confidence levels are to be *guaranteed*?

In this problem $H = .02$ and $z_q = 1.96$. An estimate of s (from the pilot survey) is

$$\begin{aligned} s &= \sqrt{\tfrac{20}{50}(1-\tfrac{20}{50})} \\ &= \sqrt{(.4)(.6)} \\ &= \sqrt{.24}. \end{aligned}$$

So

$$n = \frac{z_q^2 s^2}{H^2}$$

$$= \frac{(1.96)^2(\sqrt{.24})^2}{(.02)^2}$$

$$= 2305 \text{ students, approximately.}$$

The maximum sample size is

$$n_{\max} = \frac{.25z_q^2}{H^2}$$

$$= 2401 \text{ students.}$$

By this time you probably have noticed that one important factor—the available resources—is conspicuously absent in the sample size formulas that we discussed. However, available resources often modify the estimated sample sizes. If resources are plentiful, the sample sizes that are estimated by using the formulas should be feasible. However, if resources are limited, one may not be able to pick as large a sample as dictated by a sample size formula. For instance, if the required sample size for a study is 1500 units, but resources are available for only 1000 units, the researcher has no choice other than to pick 1000 units. Under such circumstances the sample size is dictated by the available resources, and the researcher or decision maker will simply have to accept a lower precision level (i.e., a higher value for H) or a lower confidence level (i.e., a lower value for z_q) than desired.

Another point worth noting about the sample size formula is that its application involves using data on just one study variable or question. Specifically, the estimate of the standard deviation (s) employed in the formula is for a single variable (e.g., annual expenditure on frozen foods, ownership of microwave ovens). In practice, researchers are typically interested in measuring multiple variables through the same study. When a study involves more than one variable, an estimate of the standard deviation of the most important variable (i.e., one for which the most precise estimate is desired) can be used in the formula. Alternatively, an estimate of the standard deviation for the variable expected to have the greatest variation (and have the highest standard deviation) can be used. For a given confidence level, this approach will yield a sample size that is adequate to achieve or exceed the desired precision level for all variables. When the study variables are to be measured as proportions, a sample size that will guarantee the desired precision level for all the variables can be estimated by using the formula for $n_{\max}$ given by Eq. (12.6).

Incidence Rate
In sampling, the proportion of the general population possessing a particular characteristic.

A final point to be considered in determining the sample size for some studies involving "special" populations (e.g., households with microwave ovens, families with children under five years of age) is the ***incidence rate*** of the special characteristic in question. The incidence rate of a characteristic is *the proportion of the general population possessing that characteristic.* Suppose the computed sample size for a study of households with microwave ovens is 200 and it is estimated that one-half of the households in the general population own microwave ovens (i.e., the incidence rate is .5). The actual size of the sample to be chosen from the general population so as to yield a sample of 200 households *with microwave ovens* is 200/.5 = 400 households. In general, when the incidence rate is I and the required sample size is n, the actual sample size, n', is given by

$$n' = \frac{n}{I}.$$

This adjustment to the required sample size is not needed when an adequate sampling frame of the special population is readily available, that is, if there is no need to select units from the general population and determine whether they are relevant for the study.

LIMITATIONS OF SIMPLE RANDOM SAMPLING

Although simple random sampling is the most basic probability-sampling method, it has certain drawbacks in terms of its practical usefulness. A key requirement for simple random sampling is that a complete list of population units or an exhaustive sampling frame be available. For many real-life populations—for example, the target market for perfumes, or all stores that sell paper towels and napkins—such lists may be difficult and expensive to put together. In some situations obtaining an accurate and complete sampling frame may even be impossible.

A second difficulty arises when the population and sample sizes are relatively large, which is usually the case in most practical situations. For instance, the sample selection task will be laborious if we want to randomly pick a sample of, say, 1000 units from a population of 100,000 units—unless the population list is computerized, in which case the computer can generate a simple random sample relatively easily.

A third problem with simple random sampling is that, by definition, it guarantees an equal probability of selection for every possible sample of a certain size. Hence a simple random sample can turn out to be an extreme sample. For instance, if we pick a simple random sample of two units from the hypothetical population shown in Table 12.2, the sample may well be an extreme sample, such as families 1 and 2 or families 9 and 10, which will lead to a sample statistic estimate that is very far away from the population parameter.

Because of these difficulties, and because there are other probability-sampling methods that can overcome at least some of these difficulties, simple random sampling is not very widely used in practical settings. However, our lengthy discussion of simple random sampling in this chapter does serve a useful purpose; namely, it lays the statistical groundwork that is basic to *all* probability-sampling techniques and is a foundation for more advanced probability-sampling techniques.

SUMMARY

Most marketing research projects use sampling, which is the first step in the process of making inferences about a total group of units known as a population. A population of interest to a researcher can be a collection of human or nonhuman units. A census study is one in which every population unit is examined.

When a research population is relatively large, a sampling study is likely to be less expensive, faster, and more accurate than a census study. Furthermore, given the rapidly increasing number of research surveys of human populations, sampling studies can help minimize the depletion of the pool of fresh respondents.

Census studies have a place in marketing research projects when they may be both feasible *and* necessary. The feasibility and necessity conditions for conducting a census study will be satisfied by a population that is relatively small and contains units that are distinctly different from one another.

There are two basic procedures for selecting a sample from a given population. One is probability sampling, in which the selection of individual population units is independent of a researcher's judgment or biases. The other procedure is nonprobability sampling, in which a researcher's subjective judgment does play a role in the selection of individual population units.

Simple random sampling is the most basic form of probability sampling. It involves assigning an equal probability of selection to every possible sample of a given size in a population. However, the actual selection of a simple random sample can be done by randomly picking the desired number of units from the population. This procedure is statistically equivalent to first identifying all possible samples of the desired size and then picking one of those samples at random.

Every sampling procedure has a sampling distribution associated with it. The sampling distribution shows how a sample statistic varies across random samples of a given size chosen from the same population. The nature of the distribution associated with a sampling procedure determines the extent of its sampling error. The sampling error can be quantified as the standard deviation of the sample statistic value obtained through repeated sampling; this quantified estimate is known as the standard error.

According to the central limit theorem, when the sample size is sufficiently large, the sampling distribution associated with a sampling procedure displays the properties of a normal probability distribution. This theorem allows us to construct confidence intervals for a population parameter value (mean or proportion) from a sample statistic value.

The sample size for a study must be based on the desired precision level, the desired confidence level, the degree of variability in the population to be studied, and the available resources. Formulas can be used to determine the required sample size for a study. However, since accurately estimating the standard deviation prior to conducting a study is difficult, caution must be exercised in using sample size formulas. Moreover, the sample size formulas we discussed do not directly take into account the resources available for a study. Additionally, when the focus of a study is on population units with a special characteristic, the required sample size may have to be increased by taking into account the incidence rate of the characteristic.

Simple random sampling does have some limitations. However, the principles and statistical theory underlying it form the foundation for the advanced sampling techniques to be discussed in the next chapter.

KEY TERMS

Sampling
Population (universe)
Sampling frame
Census study
Probability sampling
Nonprobability sampling
Simple random sampling
Parameter
Statistic
Sampling error
Sampling distribution
Central limit theorem
Standard error
Confidence interval
Precision level
Incidence rate

QUESTIONS

1. "When you want to pick a simple random sample of 50 units from a population of 1000 units, you do *not* need to enumerate all possible samples of 50 units." Discuss this statement.
2. What is wrong with defining *simple random sampling* as "a procedure in which each population unit has a known and equal chance of being included in the sample"? Give a more correct definition of simple random sampling.

3. Briefly discuss the distinction between *parameter* and *statistic,* using a suitable example, and define *sampling error.*

4. Define *sampling distribution.* What properties does a sampling distribution curve associated with a random-sampling procedure have? Be specific.

5. a. A simple random sample of 900 undergraduate students in Exmont University had an average grade point ratio of 2.3, with a standard deviation of 2. The sample was chosen randomly by a computer from a university record containing the names and grade point ratios of all undergraduate students in Exmont University. The sample mean and the standard deviation were also calculated and reported by the computer. Construct a 95% confidence interval for the average grade point ratio of Exmont's undergraduate population.

b. If you were asked to start from scratch and estimate the average grade point ratio *parameter,* would you use the same procedure as in part a? Or would you use a different procedure? Explain your answer. (*Hint:* Consider the pros and cons of a census study versus a sampling study.)

6. a. A marketer wants to be 95% confident that the per capita family income in a certain area is *at least* $12,000 before deciding to locate a retail outlet in the area. A simple random sample of 81 families chosen from the area had a mean income of $13,000, with a standard deviation of $4500. Should the marketer locate a retail store in the area?

b. Will your answer be different if the marketer raises the minimum per capita income requirement from $12,000 to $12,500? Why or why not?

7. a. Firm XYZ introduced a new product in a test market area. Six months after the product had been introduced, a random telephone survey of 500 households in the area indicated that 150 of them had tried the product. Construct a 95% confidence interval for the percentage of all households in the test area that had tried the new product.

b. Suppose XYZ's goal is "to ensure that at least 50% of the households in the test area try our new product within 12 months after introduction." What recommendation(s) would you make to XYZ from the results in part a?

8. Mr. J. R. Soper, a presidential candidate, does not want to campaign in any state in which he is not almost certain (i.e., 99% confident) that at least 30% of all registered voters are already in favor of him. A random telephone survey of 1000 registered voters in state A showed that 320 of them favored Mr. Soper. Should Mr. Soper campaign in state A?

9. State the four factors that should be considered in determining the sample size for any study, and indicate in what way each factor influences sample size.

10. A charitable organization is considering whether it should launch a fund-raising telethon. However, before making a "go or no-go" decision, the organization wants to conduct a national telephone survey of households to estimate the average contribution that a household will make in response to the fund-raising telethon. The organization wants to be 95% confident in the interval estimate derived from the survey, and it desires the estimate to be within ±$20. The standard deviation of charitable contributions in response to similar telethons in the past is about $125. How many households should the organization interview in its telephone survey?

11. The XYZ Company wants to conduct a study to estimate, within ±10%, the percentage of households in Texas that have microwave ovens. Moreover, XYZ wants to have 99% confidence in the interval estimate. A pilot study of 250 households in typical Texas communities found that 50 of them had microwave ovens. What sample size should XYZ use for its study?

12. A study conducted in 1980 showed that family expenditures on travel and entertainment ranged from a low of $500 to a high of $3500 per year. Although the *range* of current family expenditures on travel and entertainment is likely to be the same as in 1980, the *mean* expenditure is likely to have changed substantially. Suppose you want to estimate the current mean expenditure within ±$10, and you desire a confidence level of 95% for your estimate. How many families would you interview regarding their travel and entertainment expenditures?

13. What would your answer to Question 11 be if XYZ wanted to *guarantee* an interval estimate within ±5% at a 99% confidence level?

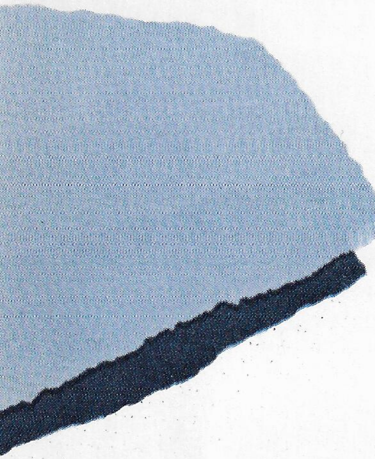
14. A firm doing a phased launch for a new product wishes to estimate the average family purchases of the new product within a precision level of 3 units and with 90% confidence. From the data of other territories, the regional manager estimates the standard deviation to be 20 units per family. Calculate the minimum sample size if the desired precision and confidence levels are to be achieved.

15. A researcher wishes to estimate the proportion of travelers visiting a gift shop at a major airport terminal. From experience in other airports, 20 out of every 100 travelers visited such stores. Estimate the required sample size if the researcher wants to be accurate within ±.05 and wants to be 95% confident of the estimate. What should be the sample size if the desired precision and confidence intervals are to be guaranteed?

16. Briefly discuss the practical constraints that limit the usefulness of formulas for determining sample size.
17. Define *incidence rate* of a characteristic and illustrate its role in sample size determination.
18. Describe the limitations of simple random sampling.

CHAPTER 13

Advanced Probability-Sampling Methods



CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Outline the steps involved in proportionate stratified random sampling.*
- *Explain why proportionate stratified random sampling is statistically more efficient than simple random sampling.*
- *Determine a confidence interval for a population mean or proportion using data obtained from a proportionate stratified random sample.*
- *State the procedural difference between proportionate and disproportionate stratified random sampling and explain why the latter has higher statistical efficiency.*
- *Outline the steps involved in simple cluster sampling.*
- *Identify the properties that clusters must have in order for simple cluster sampling to have high statistical efficiency.*
- *Compare and contrast one-stage and two-stage area sampling, and identify the trade-offs in choosing between them.*

- *State what is meant by probability-proportional-to-size area sampling and indicate when it should be used.*
- *Outline the steps involved in systematic sampling and explain why this procedure is a form of cluster sampling.*
- *Describe the potential problem due to periodicity in systematic sampling and how this problem can be overcome.*
- *Provide a general comparison of the various probability-sampling methods with respect to statistical efficiency.*
- *Identify the types of information about the population that are necessary for using each of the probability-sampling methods.*

This chapter builds on the previous one by discussing several advanced probability-sampling techniques with the aid of statistical concepts and principles developed earlier. The chapter also presents a comparison of the various probability-sampling methods and discusses their advantages and disadvantages.

STRATIFIED RANDOM SAMPLING

In simple random sampling, for a given sample size, *any* combination of population units is just as likely to become the actual sample as any other combination of the same size. Hence the actual study sample may turn out to be a skewed sample that does not adequately represent the population. The example discussed next illustrates this possibility.

Example. Suppose the administrators of some university, say Kirkwood University, want to determine the attitudes of their students toward various aspects of the university. Kirkwood University's student body, which is the population of interest here, contains 10,000 students—consisting of 3000 freshmen, 3000 sophomores, 2000 juniors, and 2000 seniors—and we want to select a sample of 500 students for conducting the attitude survey. If we use simple random sampling to pick 500 students, there is no guarantee that the chosen sample will adequately represent the four classification segments of the student body. In fact, simple random sampling may yield a sample in which all 500 students are from the *same* student classification group. In this case the sample will be extremely skewed. Such a skewed sample is not desirable since we cannot use it to generalize the survey results to the entire student body.

A sampling approach that prevents the chosen sample from being skewed is stratified random sampling.

Definition. ***Stratified random sampling*** is a probability-sampling procedure in which the chosen sample is forced to contain units from each of the segments, or strata, of the population.

Two different types of stratified random sampling are possible: *proportionate* and *disproportionate.* The next section discusses proportionate stratified random sampling and the section after that discusses disproportionate stratified random sampling.

PROPORTIONATE STRATIFIED RANDOM SAMPLING

Definition. ***Proportionate stratified random sampling*** is a form of stratified random sampling in which the sample consists of units selected from each population stratum in proportion to the total number of units in the stratum.

The following steps are involved in choosing a sample through proportionate stratified random sampling:

1. The population of interest is divided into strata on the basis of an appropriate population characteristic so that units within each stratum are alike and units across strata are different.
2. The total study sample is allocated across the population strata in such a way that the number of units allocated to each stratum is *proportionate* to the size of the stratum.
3. The allocated number of units is selected from each stratum by using a simple-random-sampling procedure.

In step 1 the research population is segmented into strata (or groups). The basis for stratifying the population is some appropriate characteristic. What kind of population characteristic will qualify as being *appropriate?* First, it must be one that is likely to be associated with the data to be collected through the research study. It must be *relevant* to the study, in the sense that one can expect the responses from the sample units to be *similar within each stratum* but *different across various strata.* Second, data about the characteristic must be readily available for each population unit *before* one conducts the study—that is, the characteristic must be *operational*—to facilitate the grouping of population units into the various strata.

The Kirkwood University example can be used to illustrate the properties of a characteristic that is appropriate for stratifying a population. Suppose we want to use proportionate stratified random sampling for selecting a sample of Kirkwood University students. Recall that the main objective of the university administrators is to ascertain students' attitudes toward the university. One characteristic that is likely to be asso-

Table 13.1 *Proportionate Allocation of Total Sample of Kirkwood University Students*

Population Strata	Number of Population Units	Number of Sample Units Allocated
Freshmen	3,000	150
Sophomores	3,000	150
Juniors	2,000	100
Seniors	2,000	100
Total	10,000	500

ciated with a student's attitudes is his or her classification—that is, whether the student is a freshman, sophomore, junior, or senior—since it reflects the nature and the scope of the student's experience at the university. For instance, the attitudes of seniors may be different from those of freshmen by virtue of the seniors' longer exposure to the university's environment. Hence classification is a relevant characteristic for the study of students' attitudes. Is it also operational? Yes, it is; the classification of each student can be identified from university records, and the student population can be easily divided into the four strata, *before* one conducts the study. Thus student classification satisfies both requirements of an appropriate characteristic.

The second step in proportionate stratified random sampling involves allocating the total sample across the population strata. This allocation is done on a proportionate basis, using the following expression:

$$\text{number of units allocated to a stratum} = \frac{\text{number of population units in the stratum}}{\text{total number of units in the population}} \times \text{total sample size.}$$

Table 13.1 illustrates the proportionate allocation of the total sample of 500 students in the Kirkwood University example. In this table, the number of units allocated to, say, the freshmen stratum is obtained as follows:

$$\frac{3{,}000}{10{,}000} \times 500 = 150$$

After the total sample is allocated, each population stratum is treated as a subpopulation, and a simple random sample of the specified number of units is chosen from each subpopulation to make up the total study sample.

In proportionate stratified random sampling the key to avoiding extreme samples is the assignment of *similar* population units to the *same* stratum and *dissimilar* population units to *different* strata. Since units that are similar will all be within the same stratum, the total sample cannot be a skewed one consisting of like units. Moreover, the selection of units from every stratum ensures adequate representation in the total sample from different segments of the population.

Sampling Distribution and Sampling Error

What impact will the procedure used in proportionate stratified random sampling have on the sampling distribution associated with it? Recall that a *sampling distribution* represents the relative frequency of occurrence of the values of a sample statistic when samples of a certain size are repeatedly chosen from the same population. If we select samples over and over again from a population by using proportionate stratified random sampling, we can expect the variation in the sample statistic value across the different samples to be less than it would be if we used a sampling procedure that allows for the possibility of skewed samples. Ensuring adequate representation of the population in any one sample is equivalent to ensuring that numerous samples chosen from the same population are as similar to one another as possible. Hence we can expect the sampling distribution of proportionate stratified random sampling to be more compact than one associated with a procedure allowing skewed samples. Therefore the *sampling error*, or *standard error*, of a proportionate-stratified-random-sampling procedure will also be relatively small.

Comparison with Simple Random Sampling

For a given sample size proportionate stratified random sampling will result in a tighter sampling distribution and a lower sampling error than will simple random sampling. To illustrate this point, let us once again consider the hypothetical population of 10 families presented in Chapter 12. In Chapter 12 we discussed the selection of a simple random sample of 2 families from this population in order to determine the mean family expenditure for eating out. Suppose we want to use proportionate stratified random sampling to select a sample of 2 families from the same population. One characteristic that will be relevant as a basis for stratifying the population is family income, since we can expect it to be associated with family expenditure for eating out. However, in real life, family income data for each population unit may not be readily available *before* one conducts a study. But for the sake of illustration, let us assume that we know whether each family in our hypothetical population is above average or below average with respect to income. Specifically, suppose families 1 through 5 have below-average incomes and families 6 through 10 have above-average incomes. Table 13.2, which is a modified version of Table 12.2, shows the stratification of the population on the basis of family income.

Table 13.2 *Stratification of a Hypothetical Population*

Population Strata	Family Number	Annual Expenditure for Eating Out ($)
Below-average income	1	50
	2	100
	3	150
	4	200
	5	250
Above-average income	6	300
	7	350
	8	400
	9	450
	10	500

For proportionate stratified random sampling the total sample of two units is allocated equally between the two strata, since they contain the same number of population units. The study sample will then consist of one unit chosen randomly from each stratum. Notice that this procedure will exclude certain sample combinations. For example, sample combinations (1, 2), (2, 4), (6, 7), and (9, 10) will not be selected through proportionate stratified random sampling. Notice also that every sample combination that is possible will have representation from each income stratum.

Table 13.3 shows the mean expenditures for eating out for the various samples of two families that can be chosen through proportionate stratified random sampling. Compare the distribution of sample mean values shown in Table 13.3 with the distribution of mean values of simple random samples of two families chosen from the same population (Table 12.2 and Exhibit 12.2 in Chapter 12). What major differences do you see between the two?

The distribution of mean values obtained through proportionate stratified random sampling differs in two distinct ways from that obtained through simple random sampling:

1. Extreme sample means do not occur for proportionate stratified random sampling. The range of possible mean values is $175 to $375 in proportionate stratified random sampling; the range is $75 to $475 in simple random sampling.
2. Sample mean values that are close to the true population mean value (i.e., the population parameter) are more likely to occur with proportionate stratified random sampling than with simple random sampling. For example, the probability of a sample mean value being

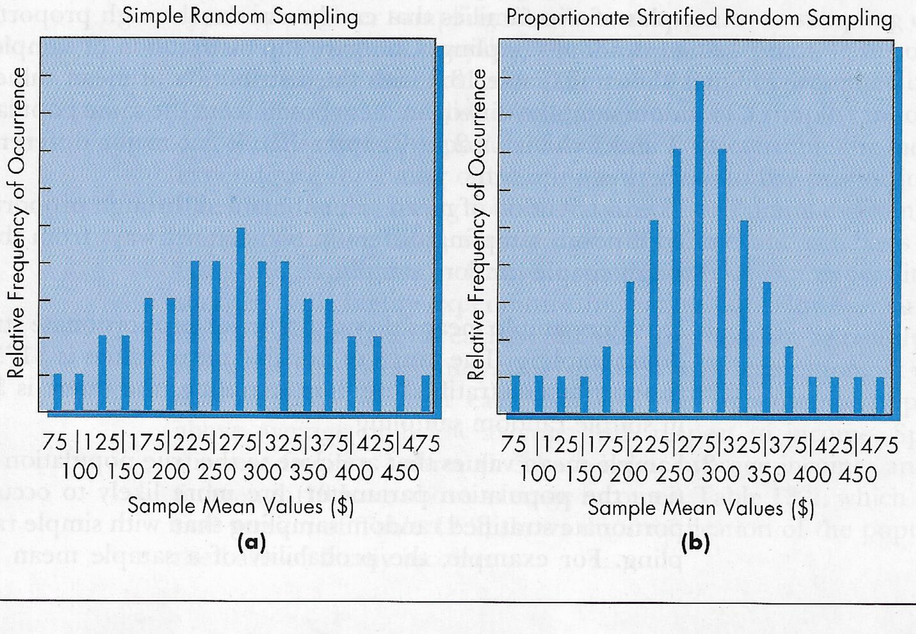
Table 13.3 *Mean Values for All Possible Proportionate Stratified Random Samples*

Possible Samples of Two Families	Sample Mean Values ($)
1, 6	175
1, 7; 2, 6	200
1, 8; 2, 7; 3, 6	225
1, 9; 2, 8; 3, 7; 4, 6	250
1, 10; 2, 9; 3, 8; 4, 7; 5, 6	275
2, 10; 3, 9; 4, 8; 5, 7	300
3, 10; 4, 9; 5, 8	325
4, 10; 5, 9	350
5, 10	375

exactly equal to the population mean value of $275 is $\frac{5}{25}$ or $\frac{1}{5}$, when proportionate stratified random sampling is used. The probability is a mere $\frac{5}{45}$ or $\frac{1}{9}$, when simple random sampling is used.

Exhibits 13.1(a) and 13.1(b) graphically demonstrate the differences between the two sampling distributions for our hypothetical population.

Exhibit 13.1 *Sampling Distributions*



(a) (b)

Exhibit 13.2 *Theoretical Sampling Distribution Curves*

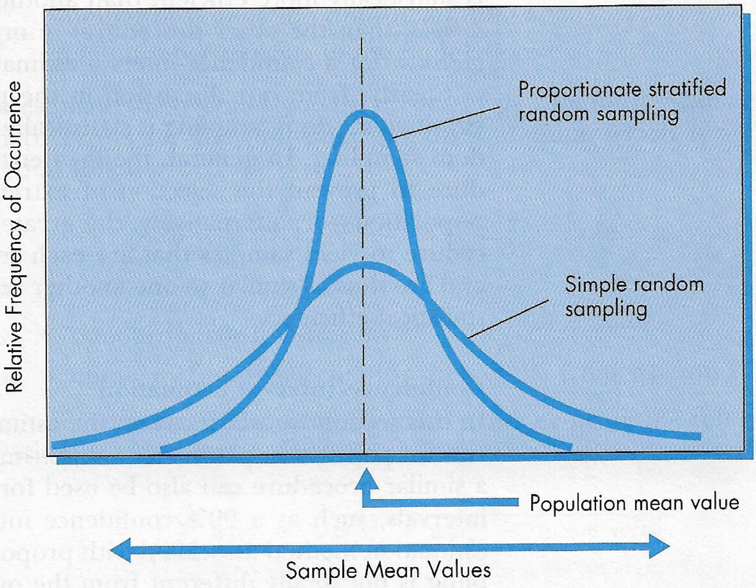


Exhibit 13.2 shows the general nature of the *theoretical* sampling distributions for simple random samples and proportionate stratified random samples of equal sizes chosen from the same population. Although both sampling distributions are centered on the population mean value, and both possess the properties of a normal probability distribution, the one associated with proportionate stratified random sampling is narrower and steeper than the one associated with simple random sampling. This result reiterates our earlier inference that for a given sample size a proportionate stratified random sample will result in a smaller sampling error (or standard error) than a simple random sample. Thus any confidence-interval estimate for the population parameter will be more precise if it is based on data obtained through a proportionate stratified random sample rather than a simple random sample.

Relative Statistical Efficiency

Definition. The ***statistical efficiency*** of a sampling procedure is a measure of how good it is in terms of the sampling error associated with it. Specifically, one sampling procedure is statistically more efficient than another procedure if for a given sample size it results in a smaller sampling error.

A sampling procedure that is statistically more efficient than another will also result in a more precise confidence-interval estimate, since the amount of sampling error and the level of precision are inversely related to each other. Alternatively, we can consider a sampling procedure that is statistically more efficient than another as one that requires a *smaller sample* (than the other procedure) in order to provide the *same level of precision* for a confidence-interval estimate.

Clearly, from our discussion in the previous section, proportionate stratified random sampling is statistically more efficient than simple random sampling. In general, the greater the ability of a sampling procedure to prevent the selection of extreme or skewed samples from a population—or, alternatively, the greater the ability of a sampling procedure to yield samples that are each representative of the population and are thus as similar to one another as possible—the higher will be its statistical efficiency.

Confidence-Interval Estimation

In this section we will focus on the estimation of 95% confidence intervals for population parameter values—means or proportions—although a similar procedure can also be used for constructing other confidence intervals, such as a 99% confidence interval. The confidence-interval estimation method associated with proportionate stratified random sampling is not totally different from the one we discussed in conjunction with simple random sampling; however, it requires some additional computations. Specifically, it involves the following steps:

1. Determine the sample statistic value (mean or proportion) and the corresponding standard error for each population stratum. Recall that any population stratum can be viewed as a subpopulation from which a simple random sample is chosen. Data obtained from each such simple random sample are used to compute a mean (or proportion) and the associated standard error.
2. Combine the sample statistic estimates for the various strata to obtain an overall statistic value.
3. Combine the standard-error estimates for the various strata to obtain an overall standard-error value.
4. Use the following expression to construct the 95% confidence interval for the population parameter:

$$\text{95\% confidence interval} = \text{overall statistic value} \pm 1.96\,(\text{overall standard error})$$

Confidence Intervals for Population Means. Let $\bar{x}_k$ be the sample mean and s_k be the corresponding standard deviation for any population stratum k. The overall sample mean, say $\bar{x}_{\text{st}}$, is estimated by combining the sample means for the various strata:

$$\bar{x}_{\text{st}} = \sum_{k=1}^{L} w_k \bar{x}_k = w_1\bar{x}_1 + w_2\bar{x}_2 + \cdots + w_L\bar{x}_L,$$

where

L = total number of strata into which the population is divided,

$$w_k = \frac{\text{number of population units in stratum } k}{\text{total number of units in the population}}$$

To estimate the overall standard error, say, $s_{\bar{x}_{\text{st}}}$, we must estimate the standard errors for the individual population strata, which is accomplished as follows:

$$s_{\bar{x}_k} = \frac{s_k}{\sqrt{n_k}} \quad \text{for} \quad k = 1, 2, \ldots, L,$$

where

$s_{\bar{x}_k}$ = standard error for stratum k,

n_k = number of units selected for the sample from stratum k.

We can now estimate the overall standard error by using the following expression:

$$s_{\bar{x}_{\text{st}}} = \sqrt{\sum_{k=1}^{L} w_k^2 s_{\bar{x}_k}^2} = \sqrt{w_1^2 s_{\bar{x}_1}^2 + w_2^2 s_{\bar{x}_2}^2 + \cdots + w_L^2 s_{\bar{x}_L}^2}.$$

The 95% confidence interval for the population mean is given by

$$\bar{x}_{\text{st}} \pm 1.96 s_{\bar{x}_{\text{st}}}.$$

Let us look at a numerical example in order to illustrate the procedure.

Problem. A professional organization consisted of a population of 3000 marketing executives, of whom 2100 had bachelor's degrees and the rest had master's degrees. A random sample of 100 executives was chosen from this population by using a proportionate-stratified-random-sampling procedure. From a survey of this sample the following results were obtained: The mean annual income of the bachelor's degree stratum was \$25,000, with a standard deviation of \$1500; the mean annual income of the master's degree stratum was \$36,000, with a standard deviation of \$2000. Determine the 95% confidence interval for the mean annual income of the members of the entire organization.

The population of interest in this problem is divided into two strata; let stratum 1 be the bachelor's degree stratum and stratum 2 be the master's degree stratum. Notice that $L = 2$ in this case. The total sample of 100 executives is allocated *proportionately* between the two strata. Therefore

$$n_1 = \frac{2100}{3000} \times 100 = 70,$$

$$n_2 = \frac{900}{3000} \times 100 = 30.$$

From the data given in the problem, we can see that $\bar{x}_1 = \$25{,}000$; $s_1 = \$1500$; $\bar{x}_2 = \$36{,}000$; and $s_2 = \$2000$. The weights for the two strata are

$$w_1 = \frac{2100}{3000} = .7,$$

$$w_2 = \frac{900}{3000} = .3.$$

The overall mean is

$$\begin{aligned}\bar{x}_{\text{st}} &= w_1\bar{x}_1 + w_2\bar{x}_2\\ &= (.7)(25{,}000) = (.3)(36{,}000)\\ &= 17{,}500 + 10{,}800 = \$28{,}300.\end{aligned}$$

The standard errors associated with the two strata are

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{1500}{\sqrt{70}},$$

$$s_{\bar{x}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{2000}{\sqrt{30}}.$$

The overall standard error is

$$\begin{aligned}s_{\bar{x}_{\text{st}}} &= \sqrt{w_1^2 s_{\bar{x}_1}^2 + w_2^2 s_{\bar{x}_2}^2}\\ &= \sqrt{(.7)^2\,\frac{(1500)^2}{70} + (.3)^2\,\frac{(2000)^2}{30}}\\ &= \sqrt{27{,}750} = \$167, \text{ approximately.}\end{aligned}$$

The 95% confidence interval is

$$\$28{,}300 \pm 1.96(\$167) = \$28{,}300 \pm \$327, \text{ approximately.}$$

From the sample data we can be 95% confident that the mean annual income of the organization's members is between \$27,973 and \$28,627.

Confidence Intervals for Population Proportions. The procedure for computing confidence intervals for proportions is the same as the one for means, except for some changes in the notation used. We will use the following notation when the variable of interest is a proportion:

$$\begin{aligned}p_k &= \text{sample proportion estimated from stratum } k,\\ s_k &= \text{standard deviation associated with stratum } k\\ &= \sqrt{p_k(1 - p_k)},\\ s_{pk} &= \text{standard error associated with stratum } k\\ &= \frac{s_k}{\sqrt{n_k}} = \sqrt{\frac{p_k\,(1 - p_k)}{n_k}},\end{aligned}$$

$$\begin{aligned} p_{\mathrm{st}} &= \text{overall proportion} \\ &= w_1 p_1 + w_2 p_2 + \cdots + w_L P_L, \end{aligned}$$

where L is the total number of strata and the w's are weights as defined in the previous section. Also,

$$\begin{aligned} s_{p_{\mathrm{st}}} &= \text{overall standard error} \\ &= \sqrt{w_1^2 s_{p_1}^2 + w_2^2 s_{p_2}^2 + \cdots + w_L^2 s_{p_L}^2}. \end{aligned}$$

The 95% confidence interval is

$$p_{\mathrm{st}} \pm 1.96 s_{p_{\mathrm{st}}}.$$

Problem. A population of interest to a marketer consists of 100,000 retail outlets. This population contains 20,000 large outlets (stratum 1), 30,000 medium outlets (stratum 2), and 50,000 small outlets (stratum 3). A proportionate stratified random sample of 1000 outlets was chosen from this population and studied. The researcher found that 60 large outlets, 150 medium outlets, and 300 small outlets in the sample were closed for business on Sundays. Determine the 95% confidence interval for the population proportion of outlets closed on Sundays.

In this problem there are three distinct strata ($L = 3$). Furthermore, the total sample of 1000 is allocated as follows:

$$n_1 = \frac{20{,}000}{100{,}000} \times 1000 = 200,$$

$$n_2 = \frac{30{,}000}{100{,}000} \times 1000 = 300,$$

$$n_3 = \frac{50{,}000}{100{,}000} \times 1000 = 500.$$

Hence $p_1 = 60/200 = .3$; $p_2 = 150/300 = .5$; and, $p_3 = 300/500 = .6$. Also,

$$w_1 = \frac{20{,}000}{100{,}000} = .2, \qquad w_2 = \frac{30{,}000}{100{,}000} = .3, \qquad w_3 = \frac{50{,}000}{100{,}000} = .5;$$

$$\begin{aligned} p_{\mathrm{st}} &= w_1 p_1 + w_2 p_2 + w_3 p_3 \\ &= (.2)(.3) + (.3)(.5) + (.5)(.6) \\ &= .06 + .15 + .30 = .51; \end{aligned}$$

$$s_{p_1} = \sqrt{\frac{p_1(1 - p_1)}{n_1}} = \sqrt{\frac{(.3)(.7)}{200}} = \sqrt{\frac{.21}{200}},$$

$$s_{p_2} = \sqrt{\frac{p_2(1 - p_2)}{n_2}} = \sqrt{\frac{(.5)(.5)}{300}} = \sqrt{\frac{.25}{300}},$$

$$s_{p_3} = \sqrt{\frac{p_3(1 - p_3)}{n_3}} = \sqrt{\frac{(.6)(.4)}{500}} = \sqrt{\frac{.24}{500}},$$

$$s_{p_{st}} = \sqrt{w_1^2 s_{p_1}^2 + w_2^2 s_{p_2}^2 + w_3^2 s_{p_3}^2}$$

$$= \sqrt{(.04)\left(\frac{.21}{200}\right) + (.09)\left(\frac{.25}{300}\right) + (.25)\left(\frac{.24}{500}\right)}$$

$$= \sqrt{.00237} = .0154, \text{ approximately.}$$

The 95% confidence interval is

$$.51 \pm 1.96(.0154) = .51 \pm .0302.$$

One can be 95% confident that the proportion of outlets closed on Sundays is between .48 and .54. Alternatively, one can be 95% confident that between 48% and 54% of the 100,000 retail outlets remain closed on Sunday.

DISPROPORTIONATE STRATIFIED RANDOM SAMPLING

In proportionate stratified random sampling the basis for allocating the total sample across the population strata is the size of each stratum. The larger the number of units in a stratum, the bigger is the sample chosen from it. However, this procedure is not necessarily the *best* way to allocate the total sample. To understand why, recall our discussion in Chapter 12 about the various factors that influence the determination of sample size. One of those factors was the *degree of variability within a population.* We saw that the greater the degree of variability, the larger the sample size should be to ensure the desired levels of confidence and precision. We can extend this principle to individual population strata, which are, after all, subpopulations. If one stratum contains more diverse units, and hence exhibits a greater degree of variability than another stratum of the same size, a larger number of units should be chosen from the former than from the latter. A sampling method that applies this principle is disproportionate stratified random sampling.

Definition. ***Disproportionate stratified random sampling*** is a form of stratified random sampling in which the sample consists of units selected from each population stratum according to how varied the units within the stratum are.

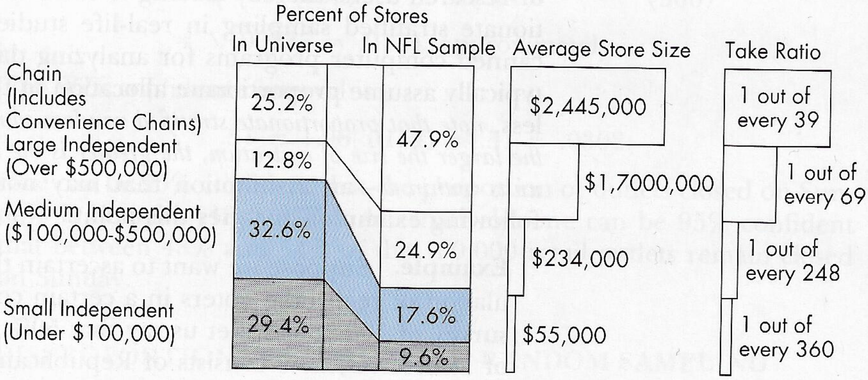
Choosing a sample through disproportionate stratified random sampling involves the same basic steps as proportionate stratified random sampling, but with one major difference in the second step: The total sample is allocated on the basis of relative variabilities, rather than sizes, of the population strata. A potential constraint in this procedure is that one may not know in advance whether the degree of diversity of the population units (with respect to the variables to be studied) differs from one stratum to another, and, if so, by how much. Consequently, in real-life situations proportionate, rather than disproportionate, stratified random sampling is more commonly used. When variances within strata are not known, though, a more sensible procedure may be to draw samples of equal rather than proportionate size from each stratum. However, proportionate samples do have a certain intuitive appeal to clients or research users, thereby leading to the relatively wider use of proportionate stratified sampling in real-life studies. Furthermore, standard canned computer programs for analyzing data from stratified samples typically assume proportionate allocation of the total sample. Nevertheless, *note that proportionate stratified random sampling implicitly assumes that the larger the size of a stratum, the greater is the degree of diversity among the units within it*—an assumption that may not be necessarily true. The following example illustrates this point.

Example. Suppose we want to ascertain the political views of a population of registered voters in a certain community by conducting a survey of 100 voters. Let us say that 50% of this population consists of Democrats, 40% consists of Republicans, and the remaining 10% consists of independents. If we were to use proportionate stratified random sampling, our sample would contain 50 Democrats, 40 Republicans, and 10 independents. However, a sample of 10 units would probably not adequately reflect the diversity of views that independents might be holding. Indeed, the diversity of political views across voters is likely to be much greater in the stratum of independents than in either the stratum of Democrats or the stratum of Republicans. Thus a better way to utilize the total sample of 100 voters is to reallocate some sample units to the stratum of independents from the other two strata. For instance, a sample consisting of 30 Democrats, 30 Republicans, and 40 independents may be more appropriate than a strictly proportionate sample.

For a given sample size disproportionate stratified random sampling will provide a better representation of the total population than proportionate stratified random sampling. This representation is accomplished by sampling *more than proportionately* (relative to their sizes) those strata that are more diverse than others. Hence disproportionate stratified random sampling will be *statistically more efficient* than proportionate stratified random sampling. If the degrees of variability across the strata are different *and known,* disproportionate, rather than proportionate, stratified random sampling must be used.

The A. C. Nielsen Company uses a form of disproportionate stratified random sampling in selecting stores that provide the data for the Nielsen Retail Index System. Exhibit 13.3 illustrates the disproportionate-sampling plan that Nielsen uses. Notice that the chain and the large-independent strata are sampled more than proportionately relative to their sizes, while the medium-independent and the small-independent strata are sampled less than proportionately relative to their sizes. Notice also that the "take ratio," which is the selection probability of a store in each stratum, varies from stratum to stratum.

Exhibit 13.3 *Disproportionate Stratified Random Sampling Used by A. C. Nielsen Company*



Source: "Management with the Nielsen Retail Index System," a brochure published by the A. C. Nielsen Company, Nielsen Plaza, Northbrook, Ill., p. 11. Courtesy of the A. C. Nielsen Company.

If proportionate sampling were used, the percentage distribution of stores across the strata in the Nielsen sample would be *identical* to the percentage distribution in the universe; moreover, the take ratio would be the same in all the four strata. The rationale for Nielsen's use of disproportionate sampling is that the chain and the large-independent strata account for the purchases of a much larger number of households and exhibit a greater variety in terms of the nature and the volume of those purchases, when compared with the other two strata. Hence the disproportionate allocation of stores is likely to provide a more accurate representation of purchases of the entire population than a proportionate allocation.

CLUSTER SAMPLING

In stratified random sampling, after partitioning the population into strata, a sample of population units is randomly selected from each stratum. This approach requires a sampling frame (i.e., list of *individual* population units) for each stratum. A sampling frame of all population units is also needed for choosing a sample through simple random sampling. An adequate sampling frame of individual population units may not always be readily available.

For example, suppose a researcher wants to conduct personal interviews with a random sample of households in a standard metropolitan statistical area (SMSA). In this scenario, a complete sampling frame of households in the SMSA may be unavailable and prohibitively expensive to compile. However, a complete listing of census tracts in the SMSA (wherein each census tract represents a cluster of households) should be readily available from the U.S. Census reports. Therefore, the researcher can randomly select a sample of *census tracts* from a sampling frame of census tracts and then interview all or some of the households in the chosen census tracts. This sampling approach is a form of cluster sampling.

Definition. ***Cluster sampling*** is a probability-sampling procedure in which *clusters* of population units are selected at random and then all or some of the units in the chosen clusters are studied.

The use of cluster sampling is not limited to situations in which a sampling frame of population units is not readily available. Even when such a sampling frame is available, if the frame can be conveniently divided into a series of representative clusters, a cluster-sampling approach may be easier to use than a simple- or stratified-random-sampling approach (an example of such a situation is presented in the next section).

Cluster sampling can take on a variety of forms, as discussed in the following sections.

SIMPLE CLUSTER SAMPLING

Definition. ***Simple cluster sampling*** (sometimes known as ***one-stage cluster sampling***) is a form of cluster sampling in which *all* population units in the selected clusters are studied.

As this definition suggests, the following steps are involved in choosing a sample through simple cluster sampling:

1. The population is divided into subpopulations, or clusters, of units so that ideally each cluster adequately represents the population.
2. A simple random sample of a few clusters is selected.
3. *All* the units in the selected clusters are studied; that is, a census of each selected cluster is conducted.

Although step 1 of simple cluster sampling resembles the first step of stratified random sampling (proportionate or disproportionate), the two sampling procedures are quite different. While the total sample in stratified random sampling consists of a sample of units from each subpopulation (stratum), the total sample in simple cluster sampling consists of *all the units within a sample of subpopulations* (clusters). Since simple cluster sampling does not involve selecting units from every cluster, its ability to provide a sample that adequately represents a population depends on the extent to which the clusters themselves are each representative of the population. Ideally, each cluster must be a miniature population;

each must adequately reflect the degree of heterogeneity (or variability) in the population and, hence, be as similar to the other clusters as possible. Thus the nature of the subpopulations formed in simple cluster sampling is quite the opposite of the nature of the subpopulations formed in stratified random sampling. While the clusters in the simple cluster sampling must be internally *heterogeneous,* and *similar* to one another, the strata in stratified random sampling must be internally *homogeneous,* and *different* from one another.

A major advantage of simple cluster sampling is the ease of sample selection relative to other probability-sampling methods. For example, suppose we have a population of 20,000 units from which we wish to select a sample of 500 units. Choosing a sample of that size through simple random sampling or stratified random sampling will require a cumbersome procedure like selecting 500 suitable numbers from a random-number table. But if the population is divided into, say, 80 clusters of 250 units each, we can select the sample relatively easily by using simple cluster sampling; we only need to pick randomly two clusters and study all the units in those chosen clusters. However, the ease of sample selection in simple cluster sampling must be weighed against the potential difficulty in forming clusters that are miniature populations. In most real-life situations a researcher will be hard-pressed to form clusters that are each heterogeneous enough to reflect the true nature of the population. The amount of work needed to form such clusters may more than offset the ease of sample selection.

Can we draw any conclusions about the relative statistical efficiency of simple cluster sampling from this discussion? Recall that the statistical efficiency of a sampling procedure depends on its ability to draw a *representative* sample from a population. Hence simple cluster sampling is likely to have relatively low statistical efficiency, unless the individual clusters are representative of the population. In fact, if the clusters are very different in composition from one another and consist of similar units within each, the statistical efficiency of simple cluster sampling may even be lower than that of simple random sampling.

To preserve the ease-of-use benefit associated with simple cluster sampling and, at the same time, make it statistically efficient, researchers must limit its practical application to research populations that can be easily divided into representative clusters. Though such populations are rare, they may be found in a few research projects, as illustrated in the following example.

Example. Consider a situation in which the sampling frame for a population consists of all the names listed in a certain city's telephone book. Let us say that there are 20,000 names in the telephone book, listed alphabetically in 200 pages, 100 names on each page. Suppose we want to choose a sample of 400 names for the research study. In this situation we can treat the 200 pages as representative clusters of the population, randomly choose any four pages, and combine all the names in those four pages to form the sample. This cluster-sampling procedure is obviously easier than other procedures such as simple random sampling. But will it also be as statistically efficient as simple random sampling? Yes, it will be if we assume that the alphabetical listing of names is as good as a random listing of names for the purpose of the research.

Although the assumption of this example may be reasonable, it may not hold up under *all* circumstances, as illustrated next.

Example. Let us say that the purpose of a research project is to study several ethnic- and culture-related aspects of people in a certain area. Moreover, suppose some large ethnic groups of people live in this area. Consequently, the last names of many individuals belonging to certain ethnic groups may all begin with the same letter. A case in point is North Orange County, California. The telephone book for this area has more Nguyen (Vietnamese) surnames than Smith surnames. Thus if the research is to be conducted in North Orange County, names listed in the telephone book cannot be considered to be a random listing. Simple cluster sampling may be inappropriate here since any sample derived through it may underrepresent or overrepresent some critical segments of the population. For instance, what if the Nguyen page in the North Orange County telephone book happens to be one of the randomly chosen clusters? The result will be a nonrepresentative sample and reduced statistical efficiency.

AREA SAMPLING

Definition. ***Area sampling*** is a form of cluster sampling in which the clusters are formed on the basis of the geographic location of the population units.

Area sampling can be *one-stage, two-stage,* or *multistage,* depending on the number of stages of random selection involved in arriving at the final sample of units. We will discuss each of these variations of area sampling, in turn.

Definition. ***One-stage area sampling*** is a form of area sampling in which *all* population units in a randomly chosen sample of subareas are studied.

Selecting a sample through one-stage area sampling is analogous to doing so through simple cluster sampling and involves the following steps:

1. Segment a total geographic area of interest (e.g., a city or a state) into subareas (e.g., city blocks, census tracts, or counties).
2. Select a few of the subareas randomly.

3. Study *all* the units (e.g., households, retail outlets, or financial institutions) in the selected subareas.

As in the case of simple cluster sampling, a key advantage of one-stage area sampling is its simplicity. It is, perhaps, even a little easier to use than simple cluster sampling since the clusters (subareas) are usually *predefined* by their geographic orientation. Unfortunately, a serious drawback of one-stage area sampling is that the final sample is not likely to be representative of the population of units. The units within any geographic subarea (e.g., households within the same city block) will be quite similar to one another on a number of characteristics. As a result, in terms of statistical efficiency, conducting a census of a few randomly chosen subareas will be a relatively poor sampling procedure. A form of area sampling that is statistically more efficient than one-stage area sampling is two-stage area sampling.

Definition. ***Two-stage area sampling*** is a form of area sampling in which randomly chosen samples of population units from each subarea of a randomly chosen sample of subareas are studied.

Two-stage area sampling differs from one-stage area sampling in the following way: Instead of the researcher's picking *all* the units from the randomly chosen subareas, only a *sample* of units is randomly picked from each of them. As its name implies, this procedure involves two distinct stages of random sampling: Stage 1 is a random selection of a sample of subareas; and stage 2 is a random selection of a sample of units from each chosen subarea. For a given sample size two-stage area sampling allows the researcher to examine a wider, more representative geographic area than one-stage area sampling; hence it will be statistically more efficient. However, on the negative side, two-stage area sampling will generally require more time and money than one-stage area sampling, owing to the wider geographic area to be covered.

Let us look at an example to illustrate the area-sampling concepts that we have discussed so far.

Example. Suppose a television station in a certain city wants to conduct a survey related to the TV-viewing behavior of households in the city. Assume that the city consists of 5000 city blocks, each of which contains 20 households. Suppose a sample of 200 households is to be chosen for the study through area sampling. Table 13.4 shows seven different combinations of city blocks (stage 1 units) and households from city blocks (stage 2 units) for obtaining a total sample of 200 households.

Plan A is a one-stage area-sampling plan; plans B through G are two-stage area-sampling plans. Notice that for a sample of 200 households plan A will be least representative of all households in the city, while plan G will be most representative. In terms of relative statistical

Table 13.4 *Alternative Sampling Plans for a Study of TV-Viewing Behavior*

Sampling Plan	Number of Stage 1 Units to be Selected	Number of Stage 2 Units to Be Selected From Each Chosen Stage 1 Unit
A	10	20
B	20	10
C	25	8
D	40	5
E	50	4
F	100	2
G	200	1

efficiency, plan A will be the worst, plan G will be the best, and the others will be in between.

How do the seven plans compare in terms of the effort and the resources needed to implement them? We can expect the ranking of the plans to be the reverse in terms of effort and resources needed for implementation: plan A best, plan G worst, and the others in between. Clearly, a choice between one-stage and two-stage area-sampling plans, or a choice from among alternative two-stage area-sampling plans, has to be based on the degree of statistical efficiency or precision desired as well as the resources available for the research.

In some research projects, multistage area sampling, which is merely an extension of two-stage area sampling, can be used.

Definition. ***Multistage area sampling*** is a form of area sampling in which more than two sequential stages of random selection are involved in the process of choosing the sample.

To illustrate multistage area sampling, suppose we want to choose a sample for a nationwide study of households in the United States. We can use multistage area sampling by randomly selecting (1) a sample of counties from a population of all counties in the United States; (2) a sample of towns and cities from each selected county; (3) a sample of census tracts from each selected town or city; and (4) a sample of households from each selected census tract. In general, for a given number of units to be selected from a specified geographic area, the greater the number of sampling stages, the greater will be the statistical efficiency and the resources required.

The methods we have discussed so far in this section illustrate the basic concepts underlying area sampling. However, they are based on some simplifying assumptions and do not exactly mirror area sampling as it is used in practice. For instance, in our treatment of two-stage area sampling we implicitly (and conveniently) assumed that each stage 1 unit had the *same* number of stage 2 units. Is this assumption valid? In many real-life situations it will not be. When the number of stage 2 units in each stage 1 unit is not the same, the stage 1 units have to be realigned or their selection probabilities have to be adjusted according to their relative sizes.

To illustrate these modifications briefly, let us again consider the example dealing with TV-viewing behavior.

Example. Suppose the 5000 city blocks making up the total area of interest in this study are not all of the same size. Specifically, suppose 2000 city blocks have 20 households each, another 2000 city blocks have 40 households each, and the remaining 1000 city blocks have 60 households each. One way of choosing an area sample of 200 households under these circumstances is to first realign the city blocks to create a new set of stage 1 units of equal size. For instance, city blocks with 40 households can each be split into two subareas with 20 households in each subarea; similarly, city blocks with 60 households can each be split into three subareas. This procedure results in a new set of 9000 stage 1 units (2000 + 2000 × 2 + 1000 × 3) with 20 households in each. The desired sample of 200 households can now be chosen by using a two-stage area-sampling plan similar to the ones in Table 13.4.

In situations where geographic realignment of stage 1 units is not feasible, an alternative approach called probability-proportional-to-size area sampling can be used.

Definition. ***Probability-proportional-to-size area sampling*** is a form of two-stage area sampling in which the probability of selection of each subarea is proportional to the number of population units in it.

In this procedure the original stage 1 units are not altered; rather, their probabilities of selection, instead of being equal, are adjusted to be proportional to the number of stage 2 units in them. In our illustration city blocks with 40 households will be twice as likely (and those with 60 households will be three times as likely) to be chosen in the first stage of sampling as city blocks with 20 households. The number of households selected in the second stage from each chosen city block will still be the same. To illustrate, suppose we want to choose 200 households by using plan E in Table 13.4 (i.e., selection of 50 city blocks followed by selection of 4 households from each). We can use probability-proportional-to-size area sampling by following these steps:

1. Number the city blocks 1 through 5000.
2. For each city block with 20 households, drop *one* piece of paper with its number written on it into a hat; drop *two* such pieces of paper into the hat for each city block with 40 households; drop *three* such pieces of paper into the hat for each city block with 60 households.

3. Draw 50 pieces of paper randomly from the hat.

4. Randomly select 4 households from each city block whose number is drawn.

The resulting sample of 200 households will provide a more accurate geographic representation of the city than a sample derived through a procedure that did not adjust the selection probabilities for the city blocks according to the number of households in them.

As the preceding modified two-stage-area-sampling illustrations demonstrate, there are several complex variations of the procedures described in this section that are frequently necessary in practical applications of area sampling. Research in Use 13.1 illustrates a multistage-area-sampling plan used by the Opinion Research Corporation. Notice that this sampling plan not only involves multiple stages but also incorporates the probability-proportional-to-size selection process.

SYSTEMATIC SAMPLING

Definition. ***Systematic sampling*** is a form of cluster sampling in which a "cluster" of units is pulled from a population by selecting the first unit randomly and the remaining units systematically.

As its name implies, systematic sampling is an organized procedure for selecting a sample from a list that contains all the population units. Specifically, it involves these steps:

Sampling Interval *The size of the gap between units chosen from a population list by using a systematic-sampling procedure.*

1. A ***sampling interval*** *k* is determined as follows:

$$k = \frac{\text{number of units in the population}}{\text{number of units desired in the sample}}.$$

2. *One unit* between the first and *k*th units in the population list is randomly chosen.

3. The randomly chosen unit and *every kth unit* thereafter are designated as part of the sample.

To illustrate the procedure, consider a sampling frame of 1000 units (numbered 1 through 1000) from which we wish to select a sample of 50 units. The sampling interval k is $1000/50 = 20$. To select the first unit for our sample, we randomly pick one number between 1 and 20, say 17. So our sample will consist of the population units numbered 17, 37, 54, . . . , 997. Notice that only the starting unit for the sample (unit 17) is randomly selected; the rest are "systematically" selected from the population list. Hence the starting unit completely determines the rest of the sample. For instance, if our initial random number happened to be 12, the units numbered 32, 52, . . . , 992 would automatically become part of the sample. Thus systematic sampling is equivalent to randomly picking *one cluster* of units (e.g., a cluster consisting of units 17, 37,

Research in Use 13.1 *Opinion Research Corporation*

Multistage Area Sampling

When studies of national scope are involved, ORC probability sampling plans are developed in the context of our National Probability Sample.

This is a permanent sampling framework which eliminates the need and expense of repeating two of the basic and most complex steps in drawing a sample in such assignments.

STEP 1: COUNTIES

The first step in developing the National Probability Sample used by Opinion Research Corporation in nationwide studies was to list the approximately 3000 counties in the United States. The list was grouped on the basis of the nine U.S. Census Bureau regions. Within each region the list was ordered by size of population.

Counties were then selected so that each county had a probability of being selected proportionate to its population.

STEP 2: TOWNS

Within each county selected, all the minor civil divisions (e.g., townships, cities) were then listed in order of population and the selection procedure repeated. That is, each town had a probability of selection proportionate to its population.

STEP 3: INTERVIEWING LOCATIONS

Once the towns were selected, starting points for the interviews were chosen. These starting points were determined from the telephone directories covering the towns selected.

STEP 4: HOUSEHOLDS

The households to be interviewed within a town were determined by selecting one or more numbers at random from the telephone book. Since some people have listed telephones while others do not, the actual interviewing starts next door to the address of the telephone number selected. In this way every household in the United States can enter into the sample.

STEP 5: INDIVIDUAL RESPONDENTS

The selection of individual respondents also is completely predesignated by the sampling plan. When the study requires that information be ob-

(Continued)

Research in Use 13.1 *Continued*

tained from a particular person in a household, such as the homemaker or the head of the house, that person is specified. In other instances, when any one of a number of persons in a household might be eligible (e.g., a study might require interviews with persons 21 and over), random methods are used to preselect the person to be interviewed.

In this situation the interviewer lists on a form all the residents of a household and their ages in a predetermined order. Each resident is assigned a number. Then the interviewer refers to a selection table to determine which person to interview.

In all cases, sound sampling techniques aim at reaching every individual included in the sample.

Source: Adapted from "Probability Sampling: A Simplified Explanation," a brochure published by the Opinion Research Corporation, 419 North Harrison Street, Princeton, NJ 08540-3574. Used by permission.

57, . . . , 997, or units 12, 32, 52, . . . , 992) from the population and considering *all* the units in that cluster as part of the sample. Conceptually, then, systematic sampling is a form of simple cluster sampling, although one need not actually divide the population into clusters if one wants to pick a systematic sample.

The most obvious attraction of systematic sampling is its simplicity relative to the other methods that we have seen: Given a population list, only one random number is required in order to select a sample from it. A less obvious benefit of systematic sampling is that in most practical applications its statistical efficiency is no lower than that of simple random sampling and can be as high as that of proportionate stratified random sampling. This benefit of systematic sampling, coupled with its simplicity, makes it very appealing from a researcher's standpoint.

The statistical efficiency of systematic sampling will vary depending on the *order* in which the population units are arranged in the list from which the sample is chosen. If the units are arranged in some random order (e.g., a population of industrial firms listed alphabetically), a systematic-sampling procedure will yield a sample that is equivalent to a simple random sample; hence its statistical efficiency will be as good as that of simple random sampling. Its statistical efficiency will be as good as that of proportionate stratified random sampling if the population units are ordered according to some relevant characteristic—one that is likely to be associated with the variables to be studied. For instance, let us say that the purpose of a research project is to estimate the average marketing expenditures of industrial firms, and that the population list of firms is ordered according to firm size, from the largest firm to the smallest, or vice versa. In this situation a systematic-sampling procedure will automatically result in proportionate representation from various size categories of industrial firms. Consequently, its statistical efficiency will be as good as that of proportionate stratified random sampling.

Systematic sampling does have one potential problem that although it arises infrequently in real life, can be quite serious when it does arise. The statistical efficiency of systematic sampling will be drastically reduced if the population list contains a certain ***periodicity*** (e.g., if every tenth unit in the list is similar) and if the periodicity happens to be equal to the sampling interval *k,* a multiple of the sampling interval (e.g., 2*k* or 3*k*), or an integral fraction of the sampling interval (e.g., *k*/2 or *k*/3). The following example illustrates this problem.

Periodicity *Any cyclical pattern with respect to a relevant characteristic in a population list used in systematic sampling. The statistical efficiency of systematic sampling will be lowered when the periodicity coincides with the sampling interval, a multiple of the sampling interval, or an integral fraction of the sampling interval.*

Example. The ABC Company, an appliance manufacturer, has a 24-hour, toll-free hot line that consumers can call if they have questions or complaints about its appliances. During the past year ABC has maintained detailed records, on an hour-by-hour basis, of the number and nature of toll-free calls it received.

Suppose ABC wants to estimate the average number of product complaint calls received per hour during the past year from each state in the United States. Moreover, suppose going through all the thousands of call records that ABC has in its files is too expensive. Hence the researcher wishes to select a *sample* of one-hour records from the chronologically arranged population of all such records.

An easy way to select a sample from the chronological listing of records is to use systematic sampling. However, there is a built-in periodicity in such a population listing: Every 24th unit will correspond to the same hour in the day. Consequently, if the sampling interval happens to be related to this built-in periodicity—sampling intervals of 12, 24, or 48, for example—a systematic sample of one-hour records will not be representative of all one-hour records; it will consist merely of records corresponding to exactly the same hour, or the same few hours, of each day. Such a sampling procedure will be statistically inefficient and lead to imprecise estimates, since we can expect the number and the origin of complaint calls to vary from hour to hour within a day.

The potential problem of periodicity in a population is not insurmountable. An alert researcher who suspects some hidden periodicity can avoid the problem through two different means. First, the researcher can use a sampling interval that is independent of the suspected periodicity—for instance, a sampling interval such as 10, 25, or 50 in the situation faced by ABC. Second, the researcher can scramble the population list to break up any periodicity and thus convert it to a random listing. However, the second approach will be cumbersome and hence may neutralize the simplicity of systematic sampling. Moreover, if maintaining a chronological or ordered list of population units is essential for other purposes, the researcher may not be able to scramble it just for the sake of using systematic sampling.

COMPARISON OF THE VARIOUS METHODS

Statistical Efficiency

Can we rank-order the various probability-sampling methods in terms of their relative statistical efficiency? To come up with a clear-cut ranking of all those methods will be somewhat difficult. The specific purpose of a research project vis-à-vis the nature of the research population involved in it (e.g., the degree of variability within the population with respect to the research questions to be addressed) can make one sampling technique statistically more efficient than another *in that research project* but not necessarily in other projects. However, we can make some generalizations about the relative statistical efficiency of different sampling methods, bearing in mind that there may be exceptions to those generalizations in a few unique situations.

In general, for a given population stratified random sampling will be statistically more efficient than simple random sampling. Disproportionate stratified random sampling will be statistically more efficient than proportionate stratified random sampling. Simple cluster sampling will be worse than simple random sampling, unless the individual clusters can be viewed as random samples from the research population. One-stage area sampling will be statistically less efficient than simple random sampling. When an area-sampling plan has two or more stages, all we can say is that its statistical efficiency will be directly related to its ability to produce a sample that adequately represents diverse geographic segments within the total area of interest. While the statistical efficiency of such a plan will be higher than that of one-stage area sampling, it may or may not be higher than that of other probability-sampling techniques. The statistical efficiency of systematic sampling will be at least equal to that of simple random sampling and can be as high as that of proportionate stratified random sampling. However, it will be lower than that of simple random sampling if the sampling interval is related to any periodicity embedded in the population list.

Practical Considerations

Not all probability-sampling methods may be equally practical in any research project. Certain factors affect the feasibility of and the effort involved in implementing them. First, unless a complete list of all population units is available, simple and stratified random sampling as well as systematic sampling cannot be used. In this regard simple cluster sampling and area sampling are somewhat more feasible because a complete list of units will be required only for the subpopulations that get selected. Second, the feasibility of some sampling plans depends on the

Table 13.5 *Comparison of Probability-Sampling Methods*

Method	Statistical Efficiency	Required Information About Population	Selection of Sample Units
Simple Random Sampling	Less efficient than stratified random sampling	List of all population units	Each unit is picked randomly
Stratified Random Sampling			
Proportionate	More efficient than simple random sampling	List of all population units; data on relevant characteristic for stratifying population	Each unit is picked randomly
Disproportionate	More efficient than simple and proportionate stratified random sampling	List of all population units; data on relevant characteristic for stratifying population; knowledge about degree of variability within each stratum	Each unit is picked randomly
Cluster Sampling			
Simple cluster	Less efficient than simple random sampling *unless* each cluster adequately represents population	List of clusters of population units	Only clusters are picked randomly (census is conducted of each chosen cluster)
Area	One-stage less efficient than simple random sampling; but two-stage more efficient than one-stage and multistage more efficient than two-stage	List of subareas for one-stage; in addition to this list, knowledge of relative sizes of the subareas, and a listing of units in the chosen subareas, for two-stage and multistage	For one-stage, only subareas are picked randomly (census is conducted of each chosen subarea); for two-stage and multistage, subareas *and* units within subareas are picked randomly

amount of prior knowledge about a population that a researcher has. For instance, both proportionate and disproportionate stratified random sampling require data on some relevant population characteristic in order for the researcher to stratify the population; additionally, disproportionate stratified random sampling requires knowledge about the degree of variability within each stratum. Two-stage area sampling, if done properly, requires knowledge about the number of stage 2 units in each stage 1 unit. Third, although every probability-sampling method involves random selection at some point, the effort needed to identify the

Table 13.5 *Continued*

Method	Statistical Efficiency	Required Information About Population	Selection of Sample Units
Systematic	As efficient as simple random sampling when population units are in random order; as efficient as proportionate stratified random sampling when population units are arranged according to a relevant population characteristic; may be less efficient than simple random sampling if there is periodicity in the population list	List of all population units	Only the first unit is picked randomly (the remaining units are picked systematically according to the sampling interval)

units for a sample varies widely across the methods. For example, in simple and stratified random sampling *each* unit must be randomly picked; in contrast, only the *first* unit needs to be randomly picked in systematic sampling.

Table 13.5 summarizes the strengths and weaknesses of the various probability-sampling methods in terms of their statistical efficiency and their feasibility as reflected by the amount of information about the population that is required and the effort involved in selecting units for the sample. Clearly, none of the methods can be considered to be the best in terms of relative statistical efficiency *and* feasibility. The choice of the most appropriate sampling technique for a research project must be based on a careful evaluation of the nature of the population involved, the degree of precision desired for the estimates, and the amount of resources available for the research.

SUMMARY

This chapter discussed several probability-sampling techniques other than simple random sampling, namely, stratified random sampling (proportionate and disproportionate), simple cluster sampling, area sampling (one-stage, two-stage, and multistage), and systematic sampling. A sizable segment of this chapter was devoted to proportionate stratified random sampling; the material there focused on extending the sampling principles and theory that we covered under simple random sampling in Chapter 12. Specifically, we discussed the nature of the sampling distribution and the estimation of confidence intervals associated with proportionate stratified random sampling vis-à-vis those associated with simple random sampling. For a given sample size proportionate stratified random sampling will have a tighter sampling distribution and will result in a more precise confidence-interval estimate than simple random sampling; that is, it will have greater statistical efficiency.

We did not explicitly examine confidence-interval estimation procedures for sampling techniques other than proportionate stratified random sampling; nor did we look at procedures for determining the required sample sizes for any of the advanced probability-sampling techniques. However, you should be aware that estimation formulas, which are somewhat more complex than the ones we have presented in Chapter 12 and in this chapter, are available for these purposes.

With respect to the relative statistical efficiency of the various sampling techniques, although some general statements can be made, an unambiguous rank ordering of all the techniques is not possible. Also, the extent to which a sampling technique is practical in a research setting will depend on factors such as the availability of a complete list of population units, the amount of prior knowledge about a population that a researcher has, and the effort needed versus the resources available to identify the sample units. Hence the choice of an appropriate technique depends to a considerable extent on the specific characteristics of and constraints in a research setting.

KEY TERMS

Stratified random sampling
Proportionate stratified random sampling
Statistical efficiency
Disproportionate stratified random sampling
Cluster sampling
Simple (one-stage) cluster sampling
Area sampling
One-stage area sampling
Two-stage area sampling
Multistage area sampling
Probability-proportional-to-size area sampling
Systematic sampling
Sampling interval
Periodicity

QUESTIONS

1. Summarize the steps involved in proportionate stratified random sampling.
2. Briefly discuss, with a suitable example, the two properties that a population characteristic must possess if it is to be appropriate for use in stratified random sampling.

3. In what ways does the sampling distribution associated with proportionate stratified random sampling differ from that associated with simple random sampling? Why?

4. Develop a logical argument to indicate why a confidence-interval estimate based on proportionate stratified random sampling will be more precise than one based on simple random sampling.

5. Define *statistical efficiency*, and indicate when one sampling procedure can be considered to be statistically more efficient than another.

6. A population of 10,000 families consists of two strata: a low-income stratum with 6400 families and a high-income stratum with 3600 families. A proportionate stratified random sample of 100 families was surveyed from this population. For the low-income stratum the mean annual expenditure for eating out was \$300, and the standard deviation was \$100. For the high-income stratum the mean was \$500, and the standard deviation was \$100. Determine the 95% confidence interval for the population mean expenditure.

7. The population of 50,000 employed adults in a city fell into the following strata: (1) 10,000 who had college degrees, (2) 15,000 who had high school diplomas but no college degrees, and (3) 25,000 who had not completed high school. A proportionate stratified random sample of 1000 individuals from this population showed the following: Those in stratum 1 had a mean income of \$20,000, with a standard deviation of \$1500. Those in stratum 2 had a mean income of \$10,000, with a standard deviation of \$1200. Those in stratum 3 had a mean income of \$8000, with a standard deviation of \$1000. What is the 95% confidence interval for the mean income of all employed adults in the city?

8. The relevant population for a study consisted of two strata: males and females. Exactly half of the population was male. A sample of 800 people was chosen from this population by using proportionate stratified random sampling. A survey of this sample indicated that 40% of the males and 20% of the females had college degrees. Estimate the 95% confidence interval for the percentage of individuals with college degrees in the population.

9. A food marketer wanted to determine the percentage of grocery stores in a certain area stocking more than three brands of catsup. There are 500 large, 1500 medium-sized, and 3000 small grocery stores in the area. A proportionate stratified random sample of 100 stores was examined. Six of the large stores, 9 of the medium-sized stores, and 15 of the small stores in the sample stocked more than three brands of catsup. Construct the 95% confidence interval for the percentage of all grocery stores stocking more than three brands of catsup.

10. A computer manufacturer just completed a sample survey of a population consisting of 4000 large businesses and 16,000 small businesses. A proportionate stratified random sample of 1000 businesses was used for the survey. The survey found that 100 large businesses in the sample and 200 small businesses in the sample had computers; the rest did not have computers. Determine the 95% confidence interval estimate for the percentage of firms with computers in the total population of firms. Give a clear interpretation of the confidence interval you computed.

11. In a certain country it was estimated that 60% of the households were in the northern half while the remaining 40% were in the southern half. A survey of a random sample of 900 households in the northern half revealed that the mean consumption of beef was 45 pounds per household per month, with a standard deviation of 30 pounds. A similar survey of 600 households in the southern half revealed a mean consumption of 35 pounds, with a standard deviation of 25 pounds. Verify that the overall sampling procedure employed in this scenario is proportionate stratified random sampling. Compute the 95% confidence interval for the mean beef consumption per household per month across the entire country.

12. What is the basic difference between proportionate and disproportionate stratified random sampling? Why does this difference make the latter procedure statistically more efficient than the former?

13. How would you describe the difference between an ideal stratum in stratified random sampling and an ideal cluster in simple cluster sampling? Explain your answer.

14. Briefly discuss the relative merits and drawbacks of simple cluster sampling.

15. Briefly discuss the pros and cons of one-stage area sampling and of two-stage area sampling.

16. List the steps involved in systematic sampling. Why is systematic sampling really a form of simple cluster sampling?

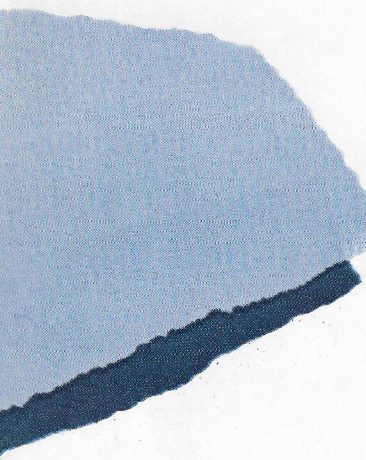
17. "The statistical efficiency of systematic sampling will never be lower than that of simple random sampling." Discuss this statement.

18. Describe the problem of periodicity in systematic sampling, and indicate ways it can be avoided.

19. What are the practical considerations that may favor one probability-sampling method over another in a research setting? Give an example to illustrate each consideration.

CHAPTER 14

Nonprobability-Sampling Methods



CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Define convenience sampling and judgment sampling and explain the difference between the two.*
- *Define implied population and ideal population.*
- *State and illustrate the ways in which the implied and ideal populations can differ when (1) convenience sampling is used, and (2) judgment sampling is used.*
- *Explain when convenience and judgment sampling procedures are appropriate.*
- *Outline the steps involved in quota sampling.*
- *Compare and contrast quota sampling with convenience sampling, judgment sampling, and stratified random sampling.*
- *State the key advantages and disadvantages of nonprobability-sampling methods over probability-sampling methods.*
- *Describe what is meant by a hybrid-sampling plan and provide an illustration of it.*

Nonprobability sampling is a subjective procedure in which the probability of selection for the population units cannot be determined. It cannot be determined because in nonprobability sampling the selection of sample units from a population is not done on a strictly chance basis, as it is in probability sampling. An important feature of nonprobability sampling is that it offers researchers greater freedom and flexibility in selecting the individual population units than does probability sampling. Before examining this feature, let us first discuss the basic types of nonprobability-sampling methods.

CONVENIENCE SAMPLING

Definition. *Convenience sampling* is a procedure in which a researcher's convenience forms the basis for selecting a sample of units.

All of us have been and continue to be exposed to numerous applications of convenience sampling during our daily routines. For instance, you have probably come across several situations similar to the ones in the following examples.

Example. The administrators of your college have just announced a sharp increase in tuition fees for the next year, citing rapidly increasing administrative costs as the main reason for the increase. This announcement is covered in the evening news program of your local TV station. The TV reporter covering this news item says, "While some of the students feel that the ten percent fee hike is justified, most of them consider it to be unfair." Then as apparent support for this conclusion, the reporter is shown standing in a certain part of the campus and talking to several students, one at a time, about their reactions to the proposed tuition fee increase.

Example. One of your marketing professors is interested in studying the impact of the size of print advertisements on the degree of awareness they are likely to create for the brands advertised. This professor has developed a set of hypothetical advertisements of different sizes for a variety of products, along with a questionnaire to measure the extent of brand awareness among consumers exposed to the advertisements. To pretest the research instrument, the professor asks the members of your class to examine the advertisements and then respond to the questionnaire.

Example. As you enter a shopping mall, you see a group of individuals passing out questionnaires to every adult entering the mall. The questionnaire is from your local government. It concerns the attitudes of local residents toward atmospheric pollution and their reactions to the possibility of increased taxes that may be needed to enforce strict pollution controls. You are asked to take just a few minutes to fill out the questionnaire and drop it in the box that is conspicuously placed in the shopping mall entrance. You notice that while a few shoppers fill out the questionnaire and drop it in the box, many others are just walking off with unfilled questionnaires.

The use of responses from a *sample* of individuals is evident in each of these examples. However, in contrast to the methods of probability sampling, notice that in each case the "researcher"—namely, the TV news reporter in the first example, the marketing professor in the second example, and the local government in the third example—was apparently willing to make some inferences on the basis of inputs from whatever sample was *conveniently available*. The researcher made no effort to locate and use inputs from a representative sample of units.

Invariably, there will be some major difference between the population as implied by a convenience sample and the ideal population.

Definition. An ***implied population*** is the collection of units to which any generalization from the actual sample is restricted.

Definition. The ***ideal population*** is the collection of units that is most pertinent for the research; it is the population to which generalizations from the research are desirable.

The implied and ideal populations in the three examples differ as shown in the list that follows.

Example	Population Implied by the Convenience Sample	Ideal Population of Interest
1. Students' reactions to proposed fee hike	Students who happened to be in the vicinity where the TV news reporter was conducting interviews on campus	All students who will have to pay higher tuition fees during the next year.
2. Pretesting research instrument	College students with at least some awareness of marketing tools, who may or may not be potential buyers of the types of products included in the test advertisements	All individuals who are likely to be potential buyers of the types of products included in the test advertisements
3. Public opinion survey about pollution	Adults who went to the shopping mall and were willing to fill out the questionnaires passed out	All adults who live within the jurisdiction of the local government

The population implied by a convenience sample can differ from the ideal population in two important ways. First, the implied population may *exclude* certain *relevant units* that are of interest to the researcher. Second, the implied population may *include* certain *irrelevant units* that are not likely to be of interest to the researcher. Illustrations of the types of relevant units excluded from, and of irrelevant units included in, the implied population in each of our three examples are given in the following list.

Example	Relevant Units Excluded	Irrelevant Units Included
1. Students' reactions to proposed fee hike	Students unwilling to be interviewed on camera Students not present on campus when the TV reporter was there	Students who will not be affected by the tuition fee increase, such as students who will graduate by next year and students who will be on tuition waiver scholarships when the fee increase takes effect

Example		Relevant Units Excluded	Irrelevant Units Included
2.	Pretesting research instrument	Nonstudent consumers with no formal education about marketing tools	Students who are not potential buyers of the advertised products
3.	Public opinion survey about pollution	Local residents who did not go to the mall when the survey was being conducted Local residents who did go to the mall but did not have time to fill out the questionnaire	Visitors who are not residents of the local area but who happened to go to the mall when the survey was being conducted

How good is a convenience sample in terms of serving the ultimate purpose of sampling, which is to be able to make some generalizations about a population (i.e., the ideal population)? The answer to this question depends on the extent of discrepancy between the ideal population and the population implied by a convenience sample. Exhibit 14.1 graphically illustrates the discrepancy between the two types of populations. The larger the shaded area in Exhibit 14.1, the more representative the convenience sample will be of the ideal population. Unfortunately, a researcher using convenience sampling has little control over how large the shaded area will be, because the researcher takes whatever respondents he or she can conveniently get. Consequently, there is no way of controlling the nature of the sample in terms of the types of respondents included in it. Indeed, in convenience sampling a researcher cannot even identify the nature of the implied population *until after such a sample has been selected and studied.*

JUDGMENT SAMPLING

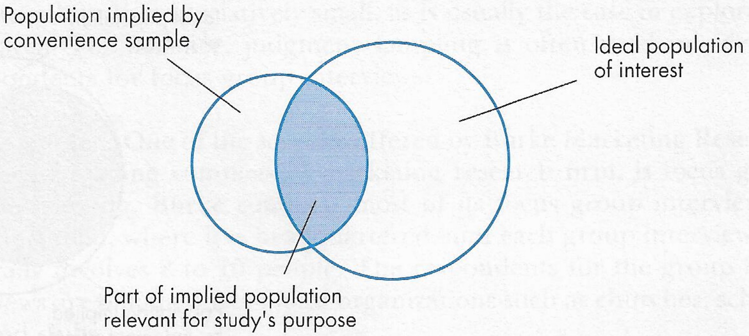
Judgment sampling is similar to convenience sampling, except that it is more refined. In convenience sampling a researcher does not exert any effort to obtain a representative sample. Such effort is expended in judgment sampling.

Definition. ***Judgment sampling*** is a procedure in which a researcher exerts some effort in selecting a sample that he or she feels is most appropriate for a study.

The definition implies that a researcher, on the basis of his or her subjective judgment, will deliberately select certain kinds of units and avoid others. For this reason judgment sampling is also known as *purposive sampling.*

Let us take another look at the TV news story example. If the TV

Exhibit 14.1 *Discrepancy Between Implied and Ideal Populations in Convenience Sampling*

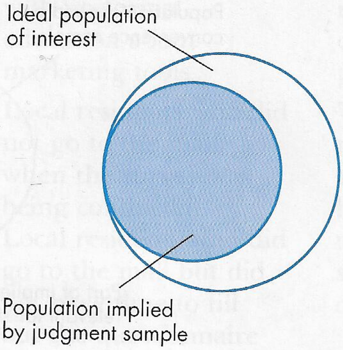


news reporter were to seek out some officers from various student organizations and obtain their reactions to the proposed fee hike, instead of merely interviewing a conveniently available sample of students, he or she would be using a judgment-sampling procedure. Whether a judgment sample is better (i.e., more representative of the ideal population) than a convenience sample will depend on how sound the judgment of the researcher is in selecting the sample units. To the extent that officers of student organizations may be well informed about the feelings of the general student body, a judgment sample of student officers may be more representative than a convenience sample of students. In practical situations the researcher will, or *should*, be knowledgeable about the nature of the relevant or ideal population for a study. Hence, although a judgment sample will require greater researcher effort, it will generally be more appropriate than a convenience sample.

Judgment sampling involves the deliberate choice of each sample unit. Assuming that the researcher is knowledgeable about the ideal population for a study and uses sound judgment in selecting the sample units, *irrelevant* units are unlikely to be included in a judgment sample. However, there is no guarantee that a judgment sample will adequately represent all *relevant* segments of the ideal population, as illustrated in the following example.

Example. Consider the judgment sample of student officers in our TV news story. Since student officers supposedly represent various student groups, the reporter may reasonably assume that they will be relevant sample units, especially if they are asked to express their perceptions of what the students in general feel about the proposed fee hike. However, can the perceptions of a sample of student officers

Exhibit 14.2 *Discrepancy Between Implied and Ideal Populations in Judgment Sampling*


Ideal population of interest
Population implied by judgment sample

capture the feelings of *all* segments of the student body? Most likely, they will not. For instance, the perceptions of a student-officer sample may not adequately reflect the views of independent students who do not belong to any student organization.

The point of this discussion is that the population implied by a judgment sample, while not likely to contain any irrelevant segment, usually will be only a *subset* of the ideal population. Exhibit 14.2 demonstrates this feature graphically.

THE USE OF CONVENIENCE AND JUDGMENT SAMPLES

Even though convenience and judgment samples are unlikely to provide perfect representations of the ideal population for a study, they can be used under certain conditions. Such samples are appropriate when the basic purpose of a research study is to generate some initial insights rather than to make any generalizations. Exploratory-research projects, which we discussed in Chapter 4, often involve the use of convenience sampling or, if time and other resources permit, judgment sampling. The next example gives an illustration.

Example. While developing Underalls—the panty hose with built-in underpants—the Hanes Company initially tested it among its women employees and obtained their reactions to it. This preliminary research suggested product benefits and promotional appeals that were further tested through additional research. The sample of women employees was clearly a convenience sample since it was most conveniently available to the Hanes management. Nevertheless, it was an appropriate sample since the purpose of the research was merely to obtain some initial reactions to the product.[1]

Judgment sampling will also be appropriate when the sample size for a research project is relatively small, as is usually the case in exploratory research. For instance, judgment sampling is often used in selecting respondents for focus group interviews.

Example. One of the services offered by Burke Marketing Research, Inc., a leading commercial marketing research firm, is focus group interviewing. Burke conducts most of its focus group interviews in Cincinnati, where it is headquartered, and each group interview typically involves 8 to 10 people. The respondents for the group interviews are obtained from local organizations such as churches, schools, and clubs.

Burke has used over 500 different organizations in the Cincinnati area to obtain respondents for its focus groups. In the judgment of Burke's researchers:

> This large number of organizations allows a degree of pre-selection of respondents. Any sample can be made *reasonably representative* [emphasis added]. Organizations tend to fall within specific age, sex, socio-economic and regional limits and a proper selection of the groups can achieve a satisfactory balance of these factors. . . . Burke gets a great deal of cooperation from the many organizations.[2]

Thus depending on the nature of the subject matter to be covered in a focus group interview, Burke selects some appropriate local organizations and requests them to provide individuals with certain desired characteristics. In short, Burke uses judgment samples for its focus group interviews.

The following important, but perhaps counterintuitive, point is implicit in this example: When a research study involves the use of a small number of sample units, a carefully chosen judgment sample may be better able to provide a proper mix of units for the study than even a probability sample.

You may find this statement hard to believe. However, consider for a moment that you are the marketing manager of a firm that wants to test-market a new product in six different metropolitan areas in the United States. You have to select the six areas from a population of all standard metropolitan statistical areas (SMSAs) in the United States. Would you select the six SMSAs on the basis of your subjective judgment, or would you prefer to select them on the basis of some probability-sampling method, say simple random sampling?

---

[1] "'Underalls' Success Due to 'Flanking Strategy,' Product Idea, Positioning," *Marketing News*, November 14, 1980, p. 11.

[2] "Group Interviewing," a brochure published by Burke Marketing Research, Inc., 2600 Victory Parkway, Cincinnati, OH 45206. Pp. 5–6.

If you use simple random sampling, there is no guarantee that the particular sample of six SMSAs chosen will be representative of the total United States market. For instance, the six SMSAs chosen by using simple random sampling may all turn out to be in California, New York, or some other limited geographic region (recall our discussion of the limitations of simple random sampling). But if you use your judgment to select the six SMSAs, you can at least attempt to balance your sample in terms of market size, geographic location, and other market characteristics that you consider relevant. Clearly, a judgment sample of six SMSAs, although not a scientifically chosen sample, may be more appropriate for test-marketing purposes than a simple random sample (or any other type of probability sample, for that matter).

QUOTA SAMPLING

Quota Sampling
A nonprobability-sampling procedure in which (1) the population is divided into cells on the basis of relevant control characteristics, (2) a quota of sample units is established for each cell, and (3) interviewers are asked to fill the quotas assigned to the various cells.

Quota sampling is the most refined form of nonprobability sampling and is often used in practice, especially in studies involving personal interviewing. Hence we will discuss it in some detail. Quota sampling resembles stratified random sampling (Chapter 13) and possesses certain features of judgment and convenience sampling as well. Quota sampling involves the following steps:

1. The population is divided into segments (typically referred to as *cells*) on the basis of certain *control* characteristics.
2. A *quota* of units to be selected from each population cell is determined from the judgment of the researchers and/or decision makers.
3. The interviewers for the study are instructed to fill the quotas assigned to the cells but are given some freedom in selecting the sample units.

A newspaper subscriber survey that the author was involved in made use of quota sampling. This survey was conducted at the request of the management of a newspaper in a medium-sized city in the Midwest. Management wanted informaton about subscribers' reactions to the newspaper and the relative popularity of various sections of it. The relevant population for this study was considered to be all adult newspaper subscribers who lived within city limits. Personal interviews were to be conducted with 300 newspaper subscribers; the sample size of 300 was based on time and cost considerations. A list of addresses of all households that subscribed to the newspaper was available from internal records.

The newspaper's management felt that reader reaction could vary depending on the sex of the reader and the section of the city in which the reader lived. Management believed that geographic location was a good surrogate indicator of the socioeconomic and educational characteristics of the city's residents. Hence sex and geographic location con-

Table 14.1 *Quota-Sampling Plan for the Newspaper Subscriber Survey*

Geographic Segment	Sex	
	Male	**Female**
I	30	30
II	30	30
III	30	30
IV	30	30
V	30	30
Total sample size = 300		

stituted the two control characteristics used in this study. The city was divided into five geographic segments, each of which was further divided into two cells on the basis of sex. As a result, ten population cells were formed. Table 14.1 shows the ten cells along with the quotas established for each of them.

Five interviewers were employed for conducting the survey, one for each geographic segment. The interviewers were given the addresses of the subscribing households located in the respective segments assigned to them. They were each asked to select 60 households and interview an adult male reader in 30 of them and an adult female reader in the remaining 30. As a safeguard against the interviewers choosing all the households from just a small section of the areas assigned to them, they were further instructed not to choose more than 4 households from any one street. Notice that even with the constraints imposed on the interviewers, they still had some freedom in selecting the households; for instance, they could, at their discretion, choose four conveniently located households (e.g., neighboring households) on any one street.

This example illustrates the general steps involved in quota sampling. Let us now examine some similarities and differences between quota sampling and other sampling methods, especially stratified random sampling. In the previous chapter we saw that the stratified-random-sampling procedure starts with dividing the population of interest into internally homogeneous strata on the basis of some appropriate population characteristic. This step resembles the first step in quota sampling: The cells and the control characteristics in quota sampling are similar to the homogeneous strata and appropriate characteristic in stratified random sampling, respectively. However, there are some major differences between quota sampling and stratified random sampling. Though each control characteristic in quota sampling must be relevant to the proposed study—that is, it must be one that is likely to be associated with the responses to the study questions—data about the characteristic need not be available for the population units *before* one conducts the study. Recall that in stratified random sampling, data on the character-

Table 14.2 *Quota-Sampling Plan for a Survey of Attitudes Toward Social Welfare Programs*

	Highest Education Level			
Age	**Less Than High School**	**High School Diploma**	**Some College**	**College Degree**
18–30	100	100	100	100
31–45	100	100	100	100
46–60	100	100	100	100
Over 60	100	100	100	100
Total sample size = 1600				

istic that serves as the basis for stratification must be available beforehand in order to facilitate the grouping of population units into the various strata and to facilitate the probabilistic selection of the designated number of units from each stratum.

In quota sampling the sample units are selected on a *subjective* rather than on a *probabilistic* basis; the only requirement of the procedure is that a designated number of units (i.e., a quota of units) possess the characteristics of each cell, irrespective of how the units are actually chosen. In quota sampling the data on control characteristics are obtained through screening questions or merely through observation (e.g., when sex is a control characteristic) *after* each potential respondent is contacted. The data are then used to determine the eligibility of each respondent for the study. The following example illustrates these features of quota sampling.

Example. Consider a study of consumer attitudes toward social welfare programs. Suppose the respondents for this study are to be selected through a quota-sampling plan in which the age and the education level of consumers are the control characteristics. The details of this quota-sampling plan are shown in Table 14.2. The first two questions in the questionnaire used in this study will be screening questions that ask for the respondent's age and education level. Each interviewer will subjectively select a respondent and ask him or her the screening questions. From the responses to the screening questions the interviewer can decide whether or not the respondent is eligible for the study. Specifically, interviews with respondents who are under 18 years of age and respondents who belong to cells whose quotas have already been filled will be terminated at this stage.

Advantages and Disadvantages

Quota sampling does not require prior knowledge about the cell to which each population unit belongs; as a result, it has two distinct advantages over stratified random sampling. First, it is easier to execute than a stratified-random-sampling procedure in which every population unit must be placed in the appropriate stratum before the actual sample selection can begin. Second, the lower a priori data requirements of quota sampling often allow a researcher to use two or more control characteristics in defining the population cells. The resultant benefit is that a researcher can control a quota sample's representativeness on several relevant population characteristics. In contrast, stratified random sampling typically, although not always, involves the use of just one characteristic as the basis for stratifying the population.

Quota sampling is superior to both convenience and judgment sampling in terms of sample representativeness. If the control characteristics used in designing a quota-sampling plan are relevant as far as the research questions are concerned, the *ideal* population of interest and the population *implied* by a quota sample will be very similar. Indeed, we can expect the ideal and the implied populations to overlap completely, at least with respect to the control characteristics.

The relative simplicity of quota sampling, when compared with probability-sampling techniques such as stratified random sampling, and its ability to provide representative samples that are much better than mere convenience or judgment samples have made it a popular technique in real-life research studies. However, quota sampling does have some potential limitations that we must be aware of. First, it lacks the statistical precision and generalizability that probability-sampling procedures possess.

Second, increasing the number of control characteristics in order to improve a sample's representativeness can sharply lower quota sampling's flexibility and make it prohibitively expensive to locate the respondents needed to fill the various cell quotas. For instance, suppose we want to add a third control characteristic to the quota-sampling plan outlined in Table 14.2. Specifically, let us say that in addition to age and education level, we want to include a respondent's income as a control characteristic with the following four levels: less than $15,000; $15,000 to less than $30,000; $30,000 to less than $50,000; and $50,000 or over. What will happen to the number of population cells in Table 14.2 if income is added as the third control characteristic? It will increase from 16 to 64. Each of the existing 16 cells will have to be further divided into 4 cells corresponding to the four levels of income. For the total sample size of 1600 respondents, the effort and the resources needed to implement a 64-cell quota-sampling plan will be substantially higher than those needed for the 16-cell quota-sampling plan. A much larger number of respondent contacts made by interviewers implementing a 64-cell quota-sampling plan may be wasted since each respondent will have to qualify for the sample on the basis of three rather than two characteristics. For instance, locating 100 respondents who are over 60 years of age and have college degrees will be relatively easier than locating 25 such respondents from each of the four income groups.

Third, even if several control characteristics are used in designing a quota sample, the sample that is actually obtained may still be biased. Quota sampling, by permitting some freedom in the selection of the actual sample units, offers interviewers an incentive to select the most convenient sample, as long as they do not violate the quota constraints. Consequently, they may make little or no effort to contact respondents who are hard to reach or difficult to talk to, such as those who live in the very rich or very poor neighborhoods. Thus a quota sample may suffer from some serious biases with respect to relevant population characteristics that are not explicitly used in forming the population cells. The real-life example described next shows that this problem can occur even in carefully designed quota samples.

Example. The National Purchase Diary Panel is maintained by NPD Research, Inc., a leading commercial marketing research firm that specializes in continuously monitoring consumer purchases of food, household products, and personal-care products. The NPD collects product purchase data from a sample of consumers who are requested to keep track of their households' purchases. The members of this diary panel are recruited by using an elaborate quota-sampling plan. Specifically, NPD has designed a 288-cell population matrix on the basis of the following five control characteristics: (1) family size, (2) age of housewife, (3) income, (4) in or out of SMSA, and (5) census division. The NPD uses the utmost care in selecting the panel members and maintaining the representativeness of its panel over time. However, by NPD's own admission:

> All panels suffer some nonrepresentivity through non-inclusion of illiterates and semi-literates, foreign speaking families, etc. . . . And, the representation of blacks is only about half of what it should be . . . despite costly efforts by NPD in recent years to increase the black population on the panel.[3]

The point is that although the population implied by a quota sample should be quite similar to the ideal population on the control characteristics, one must be cautious in generalizing the findings of a quota-sampling study. To illustrate, suppose sex, age, and education are used as control characteristics in selecting a quota sample of respondents for a study on people's attitudes toward birth control. Let us say the interviewers in this study correctly filled their quotas with interviews conducted in their own neighborhoods. But as luck would have it, all the interviewers happened to live in Catholic neighborhoods. In this case the study's findings, although based on a sample that satisfied all the specified controls, obviously cannot be generalized with a high degree of assurance.

---

[3] "Sharpening Marketing Decisions with Diary Panels," a brochure published by NPD Research, Inc., 900 West Shore Road, Port Washington, NY 11050, p. 20.

Table 14.3 *Comparison of NPD Sample with United States Population*

Family Size	Percentage of Households NPD Sample	 U.S. Census
2	37.9%	39.0%
3–4	41.4%	40.6%
5+	20.7%	20.4%
Total	100.0%	100.0%

Determining Cell Quotas

The total sample size for a study using a quota-sampling plan is usually based on how varied the population is, the depth to which various subsegments of the population are to be studied, and, most importantly, the resources available for the study. For a given total sample size, though, there is no standard procedure for dividing it into quotas for the population cells. In fact, the determination of individual-cell quotas is an area in which a researcher's judgment plays a key role. For instance, in the newspaper subscriber survey discussed earlier, the total sample of 300 respondents was allocated *equally* among the 10 population cells. In contrast, NPD uses a sophisticated scheme in setting quotas for the 288 cells represented by its diary panel. Specifically, NPD establishes quotas in such a way that on each control characteristic the distribution of panel households is approximately *proportionate* to the distribution of all households in the United States (as determined by the latest available United States census data). For example, on the dimension of family size the NPD sample and the total United States population have the percentage distributions shown in Table 14.3.

Thus researchers using quota sampling are free to establish individual-cell quotas on the basis of their best judgment about the composition of each cell and its relative importance in the overall study. As a rule of thumb, however, a quota of at least 30 units per cell is established in most real-life studies.

PROBABILITY-VERSUS NONPROBABILITY-SAMPLING METHODS

Precision

The most significant advantage of probability sampling over nonprobability sampling relates to the ability to estimate how confident one can be about a study's results. Indeed, to estimate confidence intervals for population parameters from sample data is permissible only if the sample is a probability sample. Confidence-interval estimation is not appropriate when data are gathered from a nonprobability sample. Recall that the *central limit theorem* and the theory of *normal probability distributions* form the basis for confidence-interval estimation. The application of these statistical principles is legitimate only when the sampling procedure is such that, theoretically, repeated samples chosen from the same population will yield sample statistic values that follow a *predictable* distribution, namely, a normal probability distribution. Only probability sampling, through its objectivity, can *guarantee* a process that will generate a predictable distribution pattern for the sample statistic.

To gain further insight into this advantage of probability sampling, consider that nonprobability sampling requires neither precise definition of the ideal population nor the identification of all units within it before sample selection takes place. Consequently, the only population that a nonprobability sample represents is the one it implies *after* the sample selection has taken place. Furthermore, owing to the subjectivity involved in selecting the sample units, the implied population itself could vary depending on the actual procedures and persons employed in selecting the sample. For instance, suppose a male researcher picks a judgment sample of respondents for a study of the public's opinions about sexual freedom. The population represented, or implied, by this judgment sample will most likely be different from a population implied by a judgment sample chosen by a female researcher. Furthermore, it is hard to predict in what ways and to what extent the populations implied by the two judgment samples would differ from each other and from the ideal population. Hence making any projections, in the form of confidence-interval estimates for the population parameters, from data obtained from a nonprobability sample is inappropriate and dangerous. When a nonprobability sample is used in a study, the safest approach for a researcher to take is to draw some preliminary insights after carefully examining the nature of the population implied by the particular sample chosen and the nature of the ideal population.

Time and Cost

In general, probability-sampling methods will be more time-consuming and expensive than nonprobability-sampling methods, for two reasons:

1. Probability sampling requires an accurate specification of the population and an enumeration of the units within it.
2. The selection of sample units using probability-sampling procedure must precisely follow an objective scheme that is prespecified.

Let us now examine these two reasons in greater detail. We must point out at the outset, though, that to enumerate some populations accurately may be easy. For example, a population consisting of "all dentists in Chicago" can be easily identified and listed by using the yellow pages of a telephone book or a current directory of dentists in the area. Unfortunately, in practice, research involving such simple populations is the exception rather than the rule. Populations such as "all smokers in California," "all households with children under six years of age," and "all business firms that have their own computers," which cannot be easily identified accurately, are typical of populations encountered in many research projects. Hence a researcher who wants to use probability rather than nonprobability sampling must be prepared to spend greater time and effort prior to sample selection.

Even if an accurate research population list is readily available, some time and effort will normally be needed to make contact with the *particular* units specified by the probability-sampling procedure used. For instance, suppose simple random sampling is to be used to select 100 households from a city for a personal-interview survey. The households selected through simple random sampling may be located in widely dispersed sections of the city. Consequently, contacting a simple random sample of 100 households may require greater time and expense than contacting, say, a judgment sample of 100 households. Moreover, for some households identified through simple random sampling, the interviewer may have to make several callbacks before being able to contact suitable respondents. In contrast, if some nonprobability sampling is used, there will be virtually no need for callbacks, since the interviewer does not have any objectively determined set of households to be contacted.

Determining Which Sampling Method to Use

In general, the choice between probability and nonprobability sampling involves a trade-off between the capability to generalize the sample results to the population with a *known* degree of accuracy and the lower time/cost requirements. A researcher whose primary concern is to be able to make quantitative inferences about population parameters (as is usually the case in conclusive-research projects) must use some form of probability sampling. However, time and financial constraints may force the researcher to deviate from a *rigid* probability-sampling procedure; if so, extreme caution must be used in projecting the sample results to the population. When a research project is primarily exploratory, some form of nonprobability sampling will be adequate.

Identifying the most appropriate sampling plan for a given project may not always be as clear-cut as these guidelines appear to indicate. Though nonprobability-sampling methods usually have a time and cost advantage over probability-sampling methods, the advantage may not exist in every research situation. Some sophisticated quota-sampling plans, for instance, may be just as time-consuming and expensive as probability-sampling plans. The quota-sampling scheme used by NPD in recruiting members for its diary panel is a case in point. Furthermore,

Hybrid-Sampling Plan *A sampling plan that incorporates aspects of both probability- and nonprobability-sampling methods.*

in many situations a ***hybrid-sampling plan***—one that incorporates aspects of both probability- and nonprobability-sampling methods—may be most appropriate, as illustrated by the following example.

> **Example.** Consider a study for which a representative sample of households from a metropolitan area is needed. In this study a hybrid-sampling plan, consisting of two-stage area sampling followed by quota sampling, can be used to select the desired sample of respondents. Specifically, area sampling, which only requires a list of geographic segments rather than a list of all households in the metro area, can be used initially to select a sample of census tracts and a certain number of city blocks from each selected census tract. The interviewers for the study can then be given a quota of households to be interviewed from each chosen city block. If desired, further quotas within each city block can be established with respect to other control characteristics, such as sex of the respondent and size of the household.

The point here is that depending on the objectives of the research and the resources available for conducting it, a researcher can pretty much improvise a suitable sampling plan. Such hybrid-sampling plans are used frequently in practice, as demonstrated by Research In Use 14.1, which describes the sampling plan used by the Roper Organization, Inc., in a national survey of American youth commissioned by the American Chicle group of the Warner-Lambert Company. The purpose of the survey was to ascertain the views of young people, between the ages of 8 and 17, about their perceptions of their families, their schools, and the major social issues that confront them.

However, pure probability-sampling techniques, despite their time and cost requirements, are also used in real-life projects, although perhaps less frequently. Sophisticated probability-sampling techniques are used especially by large commercial marketing research firms that maintain national samples or panels that can be readily accessed for conducting periodic research surveys. The Gallup Organization, for example, offers a business service called the Gallup Omnibus. The service is a national survey conducted every two to four weeks, using personal in-home interviews. Questions that are custom-designed to fit the requirements of individual clients can be included in the national survey in any desired month of the year. The survey's results are projectable to the United States adult civilian population 18 years and older.[4] The sample for the Gallup Omnibus is chosen through an elaborate probability-sampling plan, as described in Research In Use 14.2.

---

[4] "The Gallup Omnibus," a brochure published by the Gallup Organization, Inc., 53 Bank Street, Princeton, NJ 08540.

Research in Use 14.1 *The Roper Organization*

Area Sampling Followed by Quota Sampling

SAMPLE SIZE

A nationwide cross section of 1000 American young people, aged 8 to 17, and attending school, were interviewed for the 1987 American Chicle Youth Poll.

SAMPLING METHOD

All interviews were conducted in person in the homes of respondents and the work of all interviewers was validated by telephone. Interviews were conducted between November 15, 1986 and November 26, 1986. The sample of youth interviewed in this study is representative of the population of 8-to-17-year-olds attending school in the continental United States. A three-stage, stratified probability sample of interviewing locations was employed.

Stage one. After all the counties in the nation has been stratified by population size within each geographic region, 100 counties were selected at random proportionate to population.

Stage two. Cities and towns within the sample counties were drawn at random proportionate to population.

Stage three. Where census tract statistics within cities and towns were available, tracts were selected at random proportionate to population; where no statistics were available, rural route segments were drawn at random.

Interviewers were then given starting points at which to begin interviewing within the census tract or rural route segment. To ensure proper representation of each group in the sample, quotas were imposed for sex and age levels of respondents. While the assigned quotas produced the correct number of males and females aged 8 to 12 and 13 to 17, there were small imbalances when the sample was examined in finer age terms, e.g., too few 8- to-10-year-olds, too many 11- or 12-year-olds. Accordingly, the sample was weighted to achieve the correct proportions of males and females 8 to 10, 11 or 12, 13 or 14, and 15 to 17.

Source: Adapted from "The American Chicle Youth Poll," a report published by the Warner-Lambert Company, 201 Tabor Road, Morris Plains, NJ 07950. Used by permission.

Research in Use 14.2 *The Gallup Organization*

Sample Selection for the Gallup Omnibus

The Gallup Organization, Inc., maintains a national probability sample of 300 interviewing areas that is used for all Omnibus surveys. Each survey week, 1,500 individuals are personally interviewed. An independent sample of individuals is selected for each interviewing wave.

The sampling procedure is designed to produce an approximation of the U.S. civilian adult population living in private households, excluding those persons living in institutions such as prisons or hospitals or those living on military bases. Survey data can be applied to this population for the purpose of projecting percentages to numbers of people.

The design of the sample is that of a replicated, probability sample down to the block level in the case of urban areas, and to segments of townships in the case of rural areas. Approximately three hundred sampling points, that is, clusters of blocks or rural segments, are used in each survey. Interpenetrating samples can be provided for any given study when appropriate.

The sample design included stratification by these four size-of-community strata on the basis of Census data: Central cities of population 1,000,000 and over; 250,000 to 999,999; 50,000 to 249,999; all other populations. Each of these strata was further stratified into seven geographic regions. Within each city size—regional stratum, the population was arrayed in geographic order and zoned into equal-sized groups of sampling units. From this array of data, pairs of localities were randomly selected in each zone, with probability of selection proportional to size producing two replicated samples of localities.

Within localities so selected for which the requisite population data are reported, sub-divisions were drawn with probability of selection proportional to size of population. In all other localities, small definable geographic areas were selected with equal probability.

Separately for each survey, within each sub-division so selected for which block statistics are available, a sample of blocks is drawn with probability of selection proportional to the number of dwelling units. In all other sub-divisions or areas, blocks or segments are drawn with equal probability.

In each cluster of blocks and each segment so selected, a randomly selected starting point is designated on the interviewer's map of the area. Starting at this point, interviewers are required to follow a given direction in the selection of households until their assignment is completed.

Only one interview per household is conducted. Interviewing hours are restricted to times when adult men and women are most likely to be at home. Provision is made for a "times-at-home" weighting to allow for persons not at home when the interviewer calls.

Source: "The Gallup Omnibus," a brochure published by the Gallup Organization, Inc., 53 Bank Street, Princeton, NJ 08540. Used by permission.

SUMMARY

Convenience sampling and judgment sampling are two very basic forms of nonprobability sampling. While convenience sampling is done purely on the basis of a researcher's convenience, judgment sampling involves some researcher effort to select a representative sample. In both methods the population implied by the actual sample chosen may differ greatly and systematically from the ideal population. However, this discrepancy is likely to be somewhat less if judgment sampling is used, in the sense that a population implied by a judgment sample is not likely to contain any irrelevant units. Convenience and judgment sampling are appropriate methods when the sole purpose of a research project is to generate some initial insights. Also, judgment sampling is an ideal technique when the sample size for a study is extremely small.

Quota sampling is a refined form of nonprobability sampling that has some features of stratified random sampling, although it also involves the use of judgment in selecting the sample units. The population implied by a quota sample should be quite similar to the ideal population, at least with respect to the control characteristics. Nevertheless, a quota sample may still suffer from serious biases owing to nonrepresentativeness on dimensions that are not explicitly included as controls. The relative flexibility of quota sampling and its ability to provide reasonably representative samples have made it a popular technique. However, quota-sampling plans that involve too many control characteristics, and hence numerous population cells, may lose their advantage over probability-sampling plans with respect to ease of execution. There is no standard or universally used scheme for dividing the total sample for a quota-sampling study into individual quotas for the population cells; this area is left to the discretion of the researcher.

While nonprobability-sampling methods are, in general, less time-consuming and less expensive than probability-sampling methods, quantitative generalizations about the population on the basis of sample data, in the form of confidence-interval estimates, are permitted *only if* probability sampling is used. Hence studies using nonprobability sampling can only lead to qualitative statements, in the form of hunches or hypotheses, about the population.

The choice between probability and nonprobability sampling for a study hinges on the study's objectives as well as the resources available for conducting the study. Furthermore, hybrid-sampling plans can be devised to fit the unique circumstances surrounding a particular research project. Hybrid plans are frequently used in practice, although pure probability-sampling plans are also used in real-life studies, especially by large commercial marketing research firms.

KEY TERMS

Convenience sampling
Implied population
Ideal population
Judgment (purposive) sampling
Quota sampling
Hybrid-sampling plan

QUESTIONS

1. Define *convenience sampling*, *implied population*, and *ideal population*. In what ways can the population implied by a convenience sample differ from an ideal population? Are such differences likely to be serious?
2. Briefly discuss the differences between convenience sampling and judgment sampling.
3. Describe, with suitable examples, the circumstances in which convenience sampling and judgment sampling are appropriate.
4. List the steps involved in quota sampling. Summarize the basic procedural differences between quota sampling and stratified random sampling.
5. Discuss the pros and cons of quota sampling relative to stratified random sampling.
6. "In quota sampling the determination of individual-cell quotas is subjective." Discuss this statement.
7. Briefly discuss the advantages and limitations of nonprobability sampling relative to probability sampling.
8. Is a researcher's choice of sampling plans restricted to one of the standard probability- or nonprobability-sampling plans? Discuss your answer, giving a suitable example.
9. For each of the following situations, suggest a specific sampling plan—probability, nonprobability, or hybrid sampling—and justify your selection. Also briefly describe how the sample should be selected in each situation.
   a. A political candidate for the U.S. Senate is in a very tight race with her opponent. Election day is 10 weeks away, and this candidate wants to get a weekly reading of how much registered voters prefer her to her opponent.
   b. A large independent discount department store carries 20,000 different products. The manager of the store wants to know whether store customers are experiencing any difficulty in locating the products they are looking for.
   c. The president of an electronic-products firm wants to know what the field of computer chip technology will look like five years from now.
   d. The school district in a metropolitan area has 30 elementary schools spread across the area. The superintendent of the school district wants to ascertain the opinions of parents in the area about

the quality of elementary school education their children are receiving.

10. The management of a large shopping mall has designed a study "to determine the perceptions and likes/dislikes of the mall's customers concerning various aspects of the mall." The study design calls for a sample of 600 adult customers (300 males and 300 females) to be chosen by contacting every 10th adult customer who comes in through the mall's main entrance between 2 p.m. and 8 p.m., Monday through Saturday, of a randomly chosen week. Which of the various sampling procedures would adequately describe the proposed sampling plan? How representative is the sample likely to be of the ideal population of interest in this situation? Explain your answers.

CHAPTER 15

Data Collection: Fieldwork & Potential Errors

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Discuss the distinction between sampling error and nonsampling error.*
- *Define sampling frame error and illustrate it with an example.*
- *Define nonresponse error and discuss the potential sources of this error.*
- *Define data error and identify several forms of this error.*
- *Describe what is involved in plus-one dialing and random-digit dialing and how they help reduce sampling frame error.*
- *State ways in which nonresponse error can be reduced in mail, telephone, and personal-interview surveys.*
- *Increase the accuracy of sample means and proportions by incorporating data obtained from a subsample of nonrespondents.*
- *State several precautions that a researcher can take to reduce data error.*

The usefulness of a research project will depend to a large extent on the overall quality of the data generated. One key factor influencing data quality is sampling error. In Chapter 12 we defined ***sampling error*** as the difference between a statistic value that is generated through a sampling procedure and the population parameter value that can only be determined through an accurate census of the population. In other words, a certain amount of sampling error is bound to be present in studies that involve collecting data from only a part of the population. The higher the sampling error, the less precise (or wider) the confidence-interval estimate for the population parameter will be.

Sampling Error *A measure of the difference between a statistic value that is generated through a sampling procedure and the parameter value, which can only be determined through a census study.*

While confidence-interval estimates explicitly take into account sampling error, they do not reflect a number of other errors that can occur *irrespective* of the sampling procedure and sample size employed. These errors are called *nonsampling errors* and can arise even in census studies.

Definition. ***Nonsampling error*** is any error in a research study other than sampling error (which arises purely because a sample, rather than the entire population, is studied).

A great variety of nonsampling errors occur during the process of gathering, recording, analyzing, and interpreting data. However, nonsampling errors can be broadly classified into three basic types: sampling frame error, nonresponse error, and data error (each of these types is defined and discussed later). Strategies for reducing sampling error, such as using a statistically efficient probability-sampling technique, will have little or no effect in reducing nonsampling errors. Therefore even when sampling error in a study is quite small, the *total error* can be quite substantial if nonsampling errors are not kept in check.

Definition. ***Total error*** is the combination of sampling and nonsampling errors in a study.

The next section examines the distinction between sampling and nonsampling errors in greater detail.

SAMPLING VERSUS NONSAMPLING ERRORS

Sampling Error

Sampling error merely captures the extent of *chance*, or *random*, *fluctuations* in estimates. These fluctuations can occur from sample to sample when repeated samples (of the same size) are chosen by using the same procedure. Sampling error is only a function of the fact that a sample (rather than the entire population) is chosen for a study. It says little about the truthfulness of the collected data. The lower the variation across estimates obtained from different samples, the lower the sampling error will be.

What is the magnitude of the sampling error in a study that involves conducting a census rather than choosing a sample? It is zero, because in a census study the "sample" is the population itself. Hence repeated measures of the entire population should yield identical estimates.

In studies involving sampling the amount of sampling error can be reduced by using a sampling procedure that has high statistical efficiency and/or by increasing the sample size. Table 15.1 illustrates the magnitudes of sampling errors associated with various percentage estimates based on simple random samples of different sizes. Notice in Table 15.1 that for any given sample size the sampling error is highest for an estimate of 50%. Also notice that the sampling error associated with any percentage estimate decreases as sample size increases. Research in Use 15.1 illustrates the estimation of maximum sampling error in a *Time* magazine survey.

Nonsampling Errors

One general way of distinguishing between sampling and nonsampling errors is to say that while sampling error relates to *random variations*, which can be estimated in the form of standard error, nonsampling error typically results in some *systematic bias* that invariably is difficult, if not impossible, to estimate. Strictly speaking, some nonsampling errors may be variable errors and may cancel each other; however, such errors are assumed to be random and are thus not distinguishable from sampling error. For the sake of simplicity we will include only those errors that lead to systematic biases (i.e., definite upward or downward biases) under

Table 15.1 *Sampling Errors for Estimates of Percentages at the 95% Level of Confidence*

Sample Size	1% Or 99%	2% Or 98%	3% Or 97%	4% Or 96%	5% Or 95%	6% Or 94%	8% Or 92%	10% Or 90%	12% Or 88%	15% Or 85%	20% Or 80%	25% Or 75%	30% Or 70%	35% Or 65%	40% Or 60%	45% Or 55%	50%
25	4.0	5.6	6.8	7.8	8.7	9.5	10.8	12.0	13.0	14.3	16.0	17.3	18.3	19.1	19.6	19.8	20.0
50	2.8	4.0	4.9	5.6	6.2	6.8	7.7	8.5	9.2	10.1	11.4	12.3	13.0	13.5	13.9	14.1	14.2
75	2.3	3.2	3.9	4.5	5.0	5.5	6.2	6.9	7.5	8.2	9.2	10.0	10.5	11.0	11.3	11.4	11.5
100	2.0	2.8	3.4	3.9	4.4	4.8	5.4	6.0	6.5	7.1	8.0	8.7	9.2	9.5	9.8	9.9	10.0
150	1.6	2.3	2.8	3.2	3.6	3.9	4.4	4.9	5.3	5.9	6.6	7.1	7.5	7.8	8.0	8.1	8.2
200	1.4	2.0	2.4	2.8	3.1	3.4	3.8	4.3	4.6	5.1	5.7	6.1	6.5	6.8	7.0	7.0	7.1
250	1.2	1.8	2.2	2.5	2.7	3.0	3.4	3.8	4.1	4.5	5.0	5.5	5.8	6.0	6.2	6.2	6.3
300	1.1	1.6	2.0	2.3	2.5	2.8	3.1	3.5	3.8	4.1	4.6	5.0	5.3	5.5	5.7	5.8	5.8
400	0.99	1.4	1.7	2.0	2.2	2.4	2.7	3.0	3.3	3.6	4.0	4.3	4.6	4.8	4.9	5.0	5.0
500	0.89	1.3	1.5	1.8	2.0	2.1	2.4	2.7	2.9	3.2	3.6	3.9	4.1	4.3	4.4	4.5	4.5
600	0.81	1.1	1.4	1.6	1.8	2.0	2.2	2.5	2.7	2.9	3.3	3.6	3.8	3.9	4.0	4.1	4.1
800	0.69	0.98	1.2	1.4	1.5	1.7	1.9	2.1	2.3	2.5	2.8	3.0	3.2	3.3	3.4	3.5	3.5
1000	0.63	0.90	1.1	1.3	1.4	1.5	1.7	1.9	2.1	2.3	2.6	2.8	2.9	3.1	3.1	3.2	3.2
1200	0.57	0.81	0.99	1.1	1.3	1.4	1.6	1.7	1.9	2.1	2.3	2.5	2.7	2.8	2.8	2.9	2.9
1500	0.51	0.73	0.89	1.0	1.1	1.2	1.4	1.6	1.7	1.9	2.1	2.3	2.4	2.5	2.5	2.6	2.6
2000	0.44	0.61	0.75	0.86	0.96	1.0	1.2	1.3	1.4	1.6	1.8	1.9	2.0	2.1	2.2	2.2	2.2
2500	0.40	0.56	0.68	0.78	0.87	0.95	1.1	1.2	1.3	1.4	1.6	1.7	1.8	1.9	2.0	2.0	2.0
3000	0.36	0.51	0.62	0.71	0.79	0.87	0.99	1.1	1.2	1.3	1.5	1.6	1.7	1.7	1.8	1.8	1.8
4000	0.31	0.44	0.54	0.62	0.69	0.75	0.86	0.95	1.0	1.1	1.3	1.4	1.4	1.5	1.5	1.6	1.6
5000	0.28	0.40	0.49	0.56	0.62	0.68	0.77	0.85	0.92	1.0	1.1	1.2	1.3	1.4	1.4	1.4	1.4

Example If a sample of 500 respondents resulted in a percentage of 75%, we can be sure 95 times out of 100 that the true percentage is between 78.9% and 71.1% (75.0% ± 3.9%).

Source: Adapted from a table presented in "AIM: A Monthly Omnibus Research Service," a descriptive brochure prepared by R. H. Bruskin Associates, Market Research, 303 George Street, New Brunswick, NJ 08903, p. 35. Used by permission.

the label "nonsampling errors."[1] The following discussion illustrates the distinction between sampling and nonsampling errors.

Consider an executive club consisting of 10,000 members. You are interested in estimating the average contribution made by club members to charitable organizations on the basis of a survey of 100 members. Suppose the true average contribution (i.e., population parameter value) is $500. A probability sample of 100 club members will yield an average contribution estimate that will most likely be either lower or higher than

---

[1] For further discussion on this subject, see Leslie Kish, *Survey Sampling* (New York: Wiley, 1965), especially Chap. 13, "Biases and Nonsampling Errors," pp. 509–573. For a study showing that nonsampling error can be a significant component of total error, see Henry Assael and John Keon, "Nonsampling vs. Sampling Errors in Survey Research," *Journal of Marketing,* 46 (Spring 1982), pp. 114–123.

Research in Use 15.1 Time *Magazine Survey*

Estimation of Maximum Sampling Error

Some time ago, Yankelovich, Skelly and White, Inc., conducted a survey for *Time* magazine. According to the March 28, 1983, issue of *Time:*

> The survey polled 1,008 registered voters from March 1 to 3. The sampling error is plus or minus 3%. . . . [According to this survey] 36% of those questioned believed that cheating in income taxes is becoming more common. More unsettling, 43% found it "acceptable" to barter goods and services without reporting it on the tax forms, and 26% found it acceptable not to report cash payments as income. (p. 27)

The error of ±3% in this example is the maximum *sampling error* associated with any percentage estimate obtained from a sample of 1008 at a 95% confidence level. Let us see how the ±3% is derived. Recall from our discussion in Chapter 12 that the standard deviation, and hence standard error (which is a measure of sampling error), will be highest when $p = .5$, or 50%. Thus the maximum sampling error in the *Time* magazine example can be derived by estimating the standard error when $p = .5$ and computing the ± term associated with a 95% confidence interval:

$$\text{maximum standard error} = \sqrt{\frac{(.5)(1 - .5)}{1008}}$$

$$= \sqrt{\frac{.25}{1008}} = .0157, \text{ or } 1.57\%.$$

At the customary confidence level of 95%, a standard error of 1.57% implies a sampling error of ±(1.96)(1.57) = ±3.08%, or approximately ±3%. For percentage estimates that are greater or lower than 50%, the error will be lower than 3%. Therefore 3% represents the upper bound on sampling error associated with *any* percentage estimate from that sample.

$500. Moreover, if numerous probability samples of 100 members are chosen repeatedly, they will provide a range of average contribution estimates. However, the center of that range (or the *mean* of all the average contribution estimates) will coincide with the parameter value of $500. The mean will coincide with the parameter value because sampling error, which is the source of fluctuations across different sample estimates, is *random;* that is, high and low average contribution estimates will even out exactly and result in a mean value of $500, the true average contribution. In statistical terminology the sample average is said to be an ***unbiased estimate*** of the population average.

Unbiased Estimate
A sample statistic whose average value (over repeated samples) coincides with the population parameter it is intended to estimate.

Thus far we have implicitly assumed that each club member will *truthfully* report his or her contribution amount. Suppose this assumption is incorrect. Specifically, let us say that every club member overstates his or her contribution by $50. Under this revised assumption numerous probability samples of 100 members will still yield a range of average contribution estimates. But the mean of such estimates will *not* be $500. It will, instead, be $550; each sample average estimate will be inflated by $50. Thus the inaccurate reporting by club members will introduce a *systematic error* in the sample estimate.

A systematic error will lead to a *nonzero* net bias (either an upward or downward bias) in the sample estimate. Notice that even if the various club members overstate their contributions by *differing* amounts, there will still be an upward systematic bias in the sample estimate. Thus as long as questions exist about the accuracy of the responses provided by the club members, any estimate of their average contribution will be biased owing to nonsampling error.

Nonsampling error cannot be reduced by increasing the sample size. Indeed, when nonsampling error is present, even a census of the population will result in a biased estimate. For instance, if every club member overstates his or her contribution by $50, the average estimated contribution based on a census will be $550, in contrast to the true value of $500.

Inaccurate reporting by respondents is an illustration of one type of nonsampling error called *data error.* As mentioned earlier, there are three basic types of nonsampling error—sampling frame error, nonresponse error, and data error. These three types are discussed in the next section.

TYPES OF NONSAMPLING ERRORS

Sampling Frame Error

A *sampling frame,* as defined in Chapter 12, is a specific list of population units from which the sample for a study is chosen. For instance, suppose Sears wants to survey a sample of its credit card customers. It can readily generate a complete list of individuals who have Sears' credit cards. This list will constitute a sampling frame from which a desired number of individuals can be chosen. In this example the sampling frame is identical to the ideal population, namely, all Sears' credit card customers. In many other studies, however, generating an exact sampling frame may be quite difficult. Consequently, such studies may suffer from sampling frame error.

Definition. ***Sampling frame error*** is a bias that occurs when the population as implied by the sampling frame differs in a systematic fashion from the ideal population.

Consider a situation in which a marketer wants to conduct a telephone survey of residents of a city. An obvious sampling frame from which to choose a sample for the survey is the city's phone book. However, if we examine the appropriateness of the phone book as a sampling frame, two potential problems become evident: (1) The phone book does not contain residents with unlisted numbers, and (2) it does not contain those who do not have telephones. Given this discrepancy between the ideal population and the proposed sampling frame, a key question that arises is whether the discrepancy is serious. In other words, are the residents included in the phone book likely to differ in a systematic fashion from those not included? Yes, they are. Indeed, a research project conducted by Pacific Northwest Bell Telephone Company to examine the nature of this discrepancy gave the following conclusions:

> The results of [this project highlight] the problems inherent in using telephone directories as sampling frames. . . . In using telephone directories for sampling, systematic biases are introduced. . . . If households with unlisted numbers are systematically excluded, the resulting sample will be somewhat older, more rural, more white, more educated, more retired and more white-collar than the universe of households with telephone service.[2]

Clearly, surveys based on a sample of telephone numbers selected from a phone book will invariably contain sampling frame error. This error is a form of nonsampling error because even a census of residents with listed telephones will not yield completely accurate data. The survey results will be systematically and significantly different from those that *would* have been obtained had the sampling frame contained the entire population.

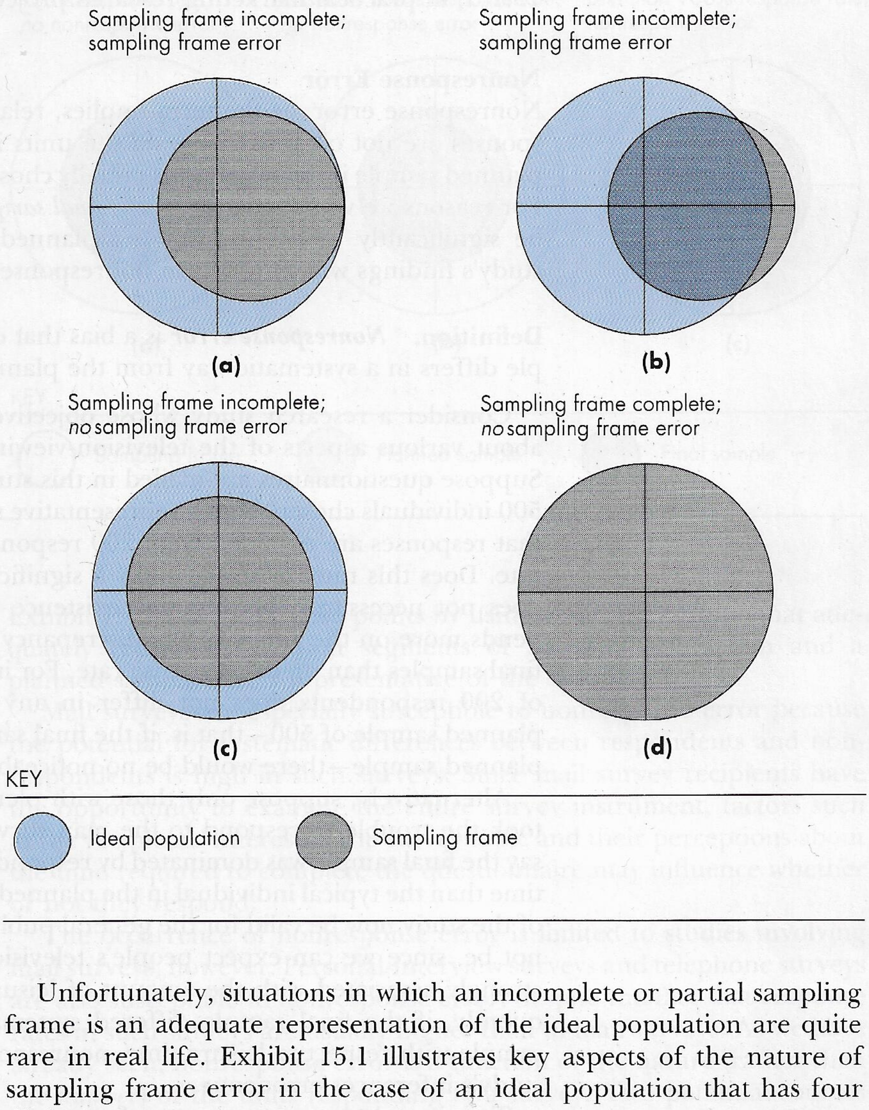
A point that needs emphasizing here, however, is that sampling frame error is not necessarily a serious problem whenever the sampling frame and the ideal population are not *identical.* In certain situations even an incomplete sampling frame may adequately represent all relevant segments of the ideal population. Consequently there may be no appreciable sampling frame error. The following example illustrates this point.

Example. The management of a retail discount store wants to survey a sample of the store's customers who use credit cards to pay for their purchases made in the store. The store has a list of names and telephone numbers of its customers who paid with credit cards during the past month. Will this list be a valid sampling frame, although it obviously does not contain *all* customers of the store who had paid for their purchases using credit cards? It most likely will be a valid sampling frame since there is no a priori reason to believe that the *composition* of customers using credit cards will vary significantly from month to month.

---

[2] Patricia E. Moberg, "Biases in Unlisted Phone Numbers," *Journal of Advertising Research,* 22 (August/September 1982), pp. 54–55.

Exhibit 15.1 *Situations Illustrating the Presence and Absence of Sampling Frame Error*


Sampling frame incomplete;
sampling frame error
(a)
Sampling frame incomplete;
sampling frame error
(b)
Sampling frame incomplete;
no sampling frame error
(c)
Sampling frame complete;
no sampling frame error
(d)
KEY
Ideal population
Sampling frame

Unfortunately, situations in which an incomplete or partial sampling frame is an adequate representation of the ideal population are quite rare in real life. Exhibit 15.1 illustrates key aspects of the nature of sampling frame error in the case of an ideal population that has four distinct segments.

Exhibit 15.1(a) shows a sampling frame that overrepresents two segments of the ideal population and underrepresents the other two. The sampling frame depicted in Exhibit 15.1(b), in addition to having a bias similar to the one in Exhibit 15.1(a), also contains units that are not included in the ideal population and hence are irrelevant. Exhibit 15.1(c) shows an incomplete sampling frame; but since it accurately represents the four population segments, it does not have sampling frame error.

In Exhibit 15.1(d) the sampling frame and ideal population are identical. Of these four situations the sampling frames in Exhibits 15.1(a) and 15.1(b) occur much more frequently than those in Exhibits 15.1(c) and 15.1(d) in practical marketing research projects.

Nonresponse Error

Nonresponse error, as the term implies, relates to studies in which responses are not obtained from all the units in the *planned sample.* The planned sample is the set of units initially chosen from a sampling frame. For reasons we will discuss later, the *final sample* obtained in a study can be significantly different from the planned sample. When it is, the study's findings will suffer from nonresponse error.

Definition. ***Nonresponse error*** is a bias that occurs when the final sample differs in a systematic way from the planned sample.

Consider a research study whose objective is to obtain information about various aspects of the television-viewing behavior of the public. Suppose questionnaires are mailed in this study to a random sample of 500 individuals chosen from a representative sampling frame. Let us say that responses are obtained from 200 respondents, for a 40% response rate. Does this rate, by itself, imply a significant nonresponse error? It does not necessarily, because the existence of nonresponse error depends more on the *nature* of the discrepancy between the planned and final samples than on the response rate. For instance, if the final sample of 200 respondents does not differ in any systematic way from the planned sample of 500—that is, if the final sample is *representative* of the planned sample—there would be no noticeable nonresponse error.

Alternatively, suppose only those with plenty of time on their hands took the trouble to respond to the mail survey. In other words, let us say the final sample was dominated by respondents who had more leisure time than the typical individual in the planned sample. Would the results of the study now be valid for the general public? They most likely would not be, since we can expect people's television-viewing behavior to be strongly associated with the amount of leisure time they have. Consequently, if the final sample differed systematically from the planned sample with respect to leisure time, nonresponse error would make the study's inferences erroneous.

Two important points emerge from our discussion of nonresponse error thus far:

1. Nonresponse error can occur even if there is no sampling frame error. Even choosing a probability sample from a representative sampling frame cannot guarantee there will be no nonresponse error.
2. The presence of nonresponse error depends on the composition of the final sample vis-à-vis the planned sample. A response rate of less than 100% does not necessarily imply there is nonresponse error.

Exhibit 15.2 *Situations Illustrating the Absence and Presence of Nonresponse Error*

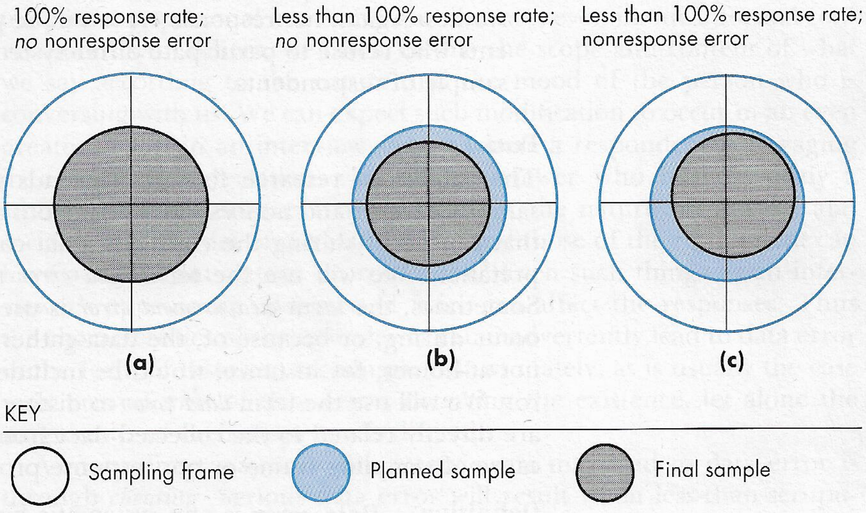


Exhibit 15.2 illustrates these points by using a sampling frame that adequately represents relevant segments of an ideal population and a planned sample that is representative of the sampling frame.

Mail surveys are especially susceptible to nonresponse error because the potential for systematic differences between respondents and nonrespondents is high in such surveys. Since mail survey recipients have the opportunity to examine the entire survey instrument, factors such as the recipients' interest in the survey topic and their perceptions about the time required to complete the questionnaire may influence whether or not they respond.

Not-at-Home Problem *A potential source of nonresponse error; it is serious when respondents who are not at home when the interviewer makes contact differ significantly from those who are at home.*

Respondent-Refusal Problem *A potential source of nonresponse error; it is serious when respondents who refuse to participate in a study differ significantly from those who agree to participate.*

The occurrence of nonresponse error is limited to studies involving mail surveys, however. Personal-interview surveys and telephone surveys are also susceptible to nonresponse error, despite the fact that response rates in such surveys are usually higher than in mail surveys. As we have already seen, nonresponse error is a function of the nature (rather than the *number*) of the units responding to a survey. Two potential sources of nonresponse error in personal and telephone surveys are as follows:

1. ***Not-at-home problem.*** Invariably, a few respondents specified by the planned sample will not be available when the interviewer attempts to make contact with them. Nonresponse error will occur if not-at-home respondents are significantly different on relevant dimensions from the respondents in the planned sample.

2. ***Respondent-refusal problem.*** Even if all the respondents in the planned sample are available when the interviewer makes contact, the possi-

bility always exists that some respondents will refuse to participate in the survey. Indeed, real-life situations in which there are *no* respondent refusals are quite rare. Refusal rates of at least 10% occur in most studies. Again, nonresponse error will be present when the respondents who refuse to participate differ systematically from the planned sample of respondents.

Data Error

The validity of research findings depends not only on the extent of sampling frame and nonresponse errors but also on the extent of errors that can occur during the course of data collection, analysis, or interpretation. We will use the term *data error* to denote any such error. Sometimes, the term *measurement error* is used to denote all errors that occur during, or because of, the data-gathering process. Errors due to not-at-homes, for instance, would be included under measurement error. We will use the term *data error* to distinguish clearly the errors that are directly related to the collected data from the errors that occur because of sampling frame or nonresponse problems.

Definition. ***Data error*** is any systematic bias that occurs during data collection, analysis, or interpretation and hence reduces the accuracy of inferences made about the ideal population.

The factors contributing to data error are so numerous that discussion of an exhaustive list is difficult. Hence we will focus on a few that are illustrative of the various types of data errors.

One potential source of data error is the respondent. In our earlier discussion of nonsampling error we saw that a systematic bias resulted when respondents inflated the magnitude of their charitable contributions. Such deliberate distortion of responses is a form of data error. Sensitive questions that invite socially acceptable answers—for example, questions about the extent to which respondents obey the law—are particularly prone to this form of data error.

Respondents may sometimes be forced to give distorted answers unintentionally. Questions that are difficult for a respondent to answer are prime candidates for distorted or erroneous responses. The following question was found in an Annual Constituent Questionnaire sent out by a U.S. congressman not too long ago:

> Do you believe people you know and associate with who receive benefits under a federal program would accept reductions in those benefits if reductions were necessary to balance the budget and make a tax cut possible and if the burden of such reductions was shared by all recipients of benefits under all federal programs?
>
> ________ Yes ________ No ________ Undecided

It is doubtful whether an average person could comprehend such a question, let alone provide a meaningful response to it. Hence the extent to which the respondents' answers to this question reflect their true feelings is open to debate.

Another potential cause of data error is the person who conducts a telephone or personal interview. For instance, even in our informal conversations with others, we often modify the scope and content of what we say according to the nature and the mood of the person who is conversing with us. We can expect such modification to occur in an even greater extent in an interview setting where a respondent is engaging in a formal conversation with an interviewer who is most likely a stranger. In fact, studies have shown that the nature of physical and social characteristics of the interviewer and those of the respondent can have a significant impact on responses.[3] Even such things as an interviewer's appearance and tone of voice can affect the responses. Thus the characteristics of the interviewer can inadvertently lead to data error in the form of distorted responses. Unfortunately, as is usually the case with many forms of data error, detecting the existence, let alone the magnitude, of such distortion is difficult.

A more direct way in which interviewers may lead to data error is through *cheating*. Serious data error will result when less-than-scrupulous interviewers even partially fill out questionnaires by themselves. However, thanks to a variety of control procedures that we will discuss later, the threat of cheating by interviewers can be reduced to a large extent.

Data errors can also occur because of sloppiness in recording, analyzing, and interpreting the data collected. For instance, mistakes may be made by interviewers in recording responses, especially those to open-ended questions. Or mistakes may be made by coders in transferring the data from questionnaires to computers. Inappropriate interpretation can also lead to data error in the form of invalid inferences. Often such errors are quite subtle, as the next example points out.

Example. The XYZ Company, the producer of brand A pain reliever, surveyed a probability sample of doctors regarding their opinions about pain relievers. One of the questions used in the survey was the following:

Would you recommend the use of brand A to your patients who suffer from aches and pains?

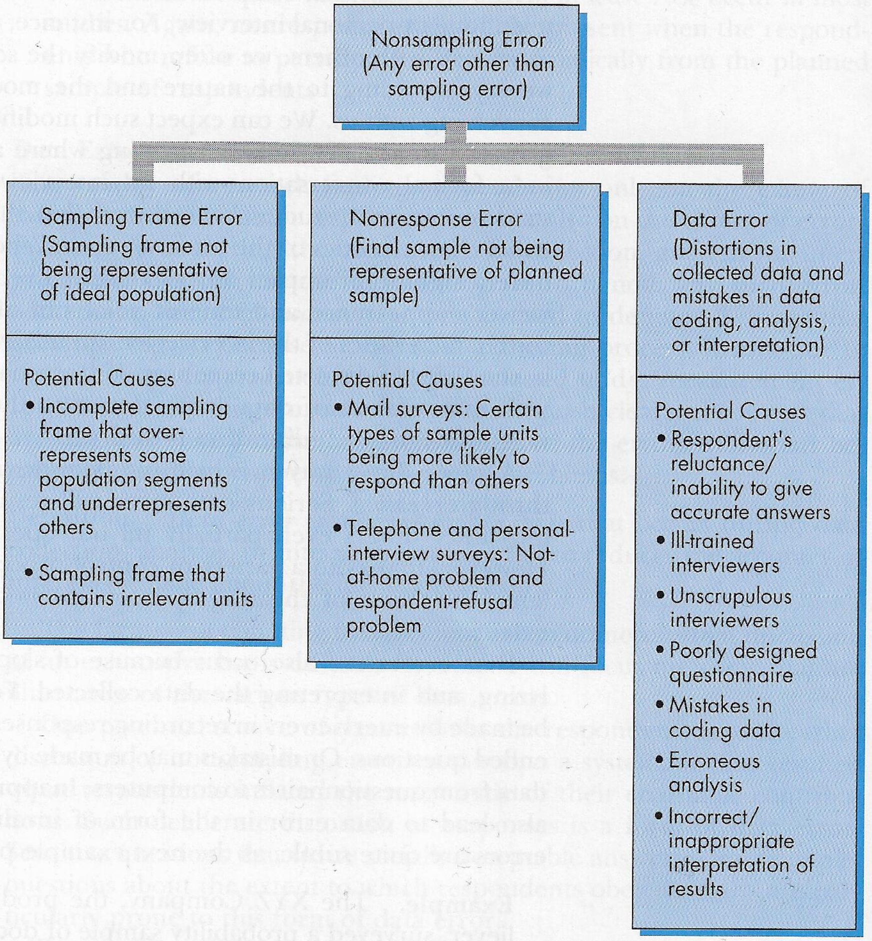
_________ Yes _________ No

Analysis of responses to this question showed that 75% of the doctors surveyed said yes; the rest said no. From this analysis XYZ

---

[3] See, for example, Robert L. Kahn and Charles L. Cannell, *The Dynamics of Interviewing* (New York: Wiley, 1957), pp. 193–196; see also U.S. Bureau of the Census, "Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960," in *Effects of Interviewers and Crew Leaders, Series ER60, no. 7* (Washington, D.C.: U.S. Government Printing Office, 1968).

Exhibit 15.3 *Types and Potential Causes of Nonsampling Errors*



Nonsampling Error (Any error other than sampling error)

Sampling Frame Error (Sampling frame not being representative of ideal population)

Potential Causes
- Incomplete sampling frame that overrepresents some population segments and underrepresents others
- Sampling frame that contains irrelevant units

Nonresponse Error (Final sample not being representative of planned sample)

Potential Causes
- Mail surveys: Certain types of sample units being more likely to respond than others
- Telephone and personal-interview surveys: Not-at-home problem and respondent-refusal problem

Data Error (Distortions in collected data and mistakes in data coding, analysis, or interpretation)

Potential Causes
- Respondent's reluctance/inability to give accurate answers
- Ill-trained interviewers
- Unscrupulous interviewers
- Poorly designed questionnaire
- Mistakes in coding data
- Erroneous analysis
- Incorrect/inappropriate interpretation of results

Company made the following inference. "Three out of four doctors surveyed recommend brand A for their patients who suffer from aches and pains."

Is the inference made by XYZ Company valid? At first glance it appears to be. However, some doubt arises about its meaning, and hence accuracy, when one critically examines the *wording* of the question used in the survey. For instance, consider the following alternative wording for the same question:

> Which brand of pain reliever would you recommend the most to your patients who suffer from aches and pains?

__________ Brand A

__________ Brand B

__________ Brand C

.
.
.

__________ Other ____________________ (Please specify.)

The alternative wording obviously offers greater latitude of response for the doctors than the wording used in the survey, which forced the doctors to say yes or no to brand A. Consequently, one can argue that the only valid inference that can be drawn from the survey's findings is something like this: "Three out of four doctors surveyed *do not object* to the use of brand A by their patients who suffer from aches and pain." Alternatively, unless 75% of the doctors checked brand A in response to the revised question, one could not legitimately make the inference that XYZ Company made. The point here is that the danger of some form of data error exists even long after the data is collected and coded.

In summary, of the two components of *total error* in a research project, only sampling error can be influenced by the choice of sample size and the sample selection procedure to be used. Nonsampling errors—consisting of sampling frame error, nonresponse error, and data error—cannot be lowered by merely using large sample sizes and/or statistically efficient sample selection procedures. Exhibit 15.3 summarizes the three types of nonsampling errors and their potential causes. The next section discusses ways of controlling these errors.

CONTROLLING NONSAMPLING ERRORS

Sampling Frame Error

The only way to guarantee the absence of sampling frame error is to have a *complete* sampling frame when you begin a study. However, developing a complete sampling frame may often be tedious or impossible. Consider the situation presented in the following example.

Example Home Sales Company, a real estate brokerage firm, wants to conduct a personal-interview survey of a sample of people who have bought their homes within the past year. Home Sales, as well as all other realtors in its area of operation, subscribe to a service known as the Multiple-Listing Service, which has a list of all buyers who purchased homes listed through local realtors during the past year. Home Sales has ready access to this list.

Will this list adequately represent all those who bought homes within the past year? It will be a complete sampling frame only if the population of interest is defined as the buyers of homes *listed through realtors*. In other words, missing from this list are all those who bought their homes directly from the owners. Admittedly, direct buyers are usually only a small fraction of all home buyers. However, their characteristics and decision-making behavior are likely to be significantly different from those who bought their homes through realtors. Hence excluding them can lead to serious sampling frame error if the ideal population consists of *all* home buyers.

The only way to overcome this error is to develop a complete sampling frame, perhaps by going through the records of the tax offices in the area. This process will be more cumbersome than using the readily available sampling frame, but it is necessary if sampling frame error is to be avoided.

Plus-One Dialing *A method of reducing sampling frame error in telephone surveys; it involves adding a fixed integer (usually 1) to each telephone number chosen from a phone book so as to obtain a revised number to be dialed.*

Sampling frame error can also be reduced in some situations by modifying an available sampling frame that is incomplete. For instance, we saw earlier that a phone book is invariably an incomplete sampling frame when the ideal population is the general public. One way of overcoming this problem, at least partially, is to use a technique called ***plus-one dialing***. In plus-one dialing a fixed integer (usually 1) is added to each telephone number chosen from the phone book, and the revised telephone number is dialed. In theory, plus-one dialing amounts to using a sample of numbers chosen from a *modified* sampling frame, which is constructed by increasing each telephone number in the original sampling frame by a fixed number. The rationale behind plus-one dialing is that the modified sampling frame will constitute an adequate cross section of listed and unlisted numbers, thus leading to a reduction in sampling frame error.

Random-Digit Dialing *A method of reducing sampling frame error in telephone surveys; it involves randomly dialing the last four digits of telephone numbers within any exchange specified by a three-digit prefix.*

Another technique used to reduce sampling frame error in telephone surveys is ***random-digit dialing***. This technique involves randomly dialing the last four digits of telephone numbers within each telephone exchange included in the area of interest. For instance, a random sample of telephone numbers can be obtained from any exchange (say, 555) by randomly generating the X's in 555–XXXX and avoiding any duplication of numbers already chosen. With the advent of high-speed computers and computerized telephone dialing, random-digit dialing is easy and quick to use if one has the appropriate equipment. Sievers Research Company, for example, uses random-digit dialing in its Omniview services, which is a regular monthly survey of a national probability sample of adults.[4]

Neither plus-one dialing nor random-digit dialing can completely eliminate sampling frame error when the ideal population is *all* households—those with and those without telephones. That is, any telephone survey, irrespective of the technique used to select the numbers, can only represent households that have telephone service.

---

[4] "Check on America with OMNIVIEW," a brochure prepared by Sievers Research Company, 2111 Huntington Drive, San Marino, Calif. 91108.

Nonresponse Errors

Mail Surveys. Nonresponse errors in mail surveys can be reduced by increasing the response rate in such a way as to make the final sample reasonably representative of the planned sample. A number of approaches have been found useful in increasing response rates. For example, several studies have shown the following techniques effective in this regard: the inclusion of cash or other incentives with the questionnaire; the mailing of follow-up questionnaires or reminders; and the use of first-rate postage (e.g., first-class mail, special delivery) to mail the questionnaire.[5]

Of course, employing any of these techniques adds to the cost of conducting a mail survey. Moreover, the available research evidence is not clear about the impact of these response rate improvement techniques on the *composition* of respondents in the final sample. A potential danger is that some of these techniques may increase the size, but not necessarily the representativeness, of the final sample, as indicated in the next example.

Example. The National Purchase Dairy Panel of NPD Research, Inc., contains families who are requested to keep a diary of their purchases and mail the diaries to NPD at regular intervals. The NPD uses *gifts* (rather than cash or cashlike items such as coupons or trading stamps) to encourage panel families to make entries in their diaries and mail them in regularly. The following rationale provided by NPD for using gifts as incentives illustrates the "potential danger" alluded to earlier:

> Compensating families via gifts is superior to using trading stamps, although stamps are less expensive and this method is used by one competitive panel. Our research shows stamp savers are atypical of the population (greater price sensitivity) and their choice of stores is obviously influenced by the availability of stamps.[6]

Thus nonresponse error may be a problem in mail surveys even if the response rate is adequate. Hence if at all possible, one should compare the composition of the planned sample with that of the final sample—or, alternatively, the composition of respondents with that of nonrespondents—on characteristics that are relevant to the study. If the comparison reveals substantial differences between the two groups, the researcher should exercise caution in interpreting the results of the study, owing to the possible presence of nonresponse error.

---

[5] Leslie Kanuk and Conrad Berenson, "Mail Surveys and Response Rates: A Literature Review," *Journal of Marketing Research*, 12 (November 1975), pp. 440–53; Arnold S. Linsky, "Stimulating Responses to Mailed Questionnaires: A Review," *Public Opinion Quarterly*, Spring 1975, pp. 82–101; Don A. Dillman, *Mail and Telephone Surveys* (New York: Wiley, 1978).

[6] "Sharpening Marketing Decisions with Diary Panels," a brochure prepared (in 1975) by NPD Research, Inc., 900 West Shore Road, Port Washington, NY 11050, p. 22.

The author was involved in a research project aimed at generating information about new-product-planning activities of manufacturing firms in the United States. A random sample of 500 manufacturing firms was chosen, and a questionnaire (along with a cover letter) was mailed to the president of each firm. The sample of 500 firms was the planned sample. A second questionnaire (follow-up) was mailed to those firms that had not responded within five weeks. A total of 107 firms returned properly filled-out questionnaires; this sample was the final sample. Nonresponse error was a concern in this study, especially since the effective response rate was only a little over 21%. Fortunately, data were available for both the planned and the final sample on certain key characteristics that were likely to be associated with new-product-planning activities. One such characteristic was the type of industry to which each firm belonged. Table 15.2 offers a comparison of the firms in the planned and final samples on this characteristic.[7]

Although there are differences in the composition of the two samples (as indicated by the percentage figures in Table 15.2), they do not appear to be major. Moreover, these differences were not statistically significant on the basis of a chi-square test (we will discuss chi-square tests in Chapter 17). Since the planned and final samples were similar in composition, we concluded that nonresponse error was negligible.

Telephone and Personal-Interview Surveys. As we discussed earlier, two primary sources of nonresponse error in telephone and personal-interview surveys are the not-at-home and respondent-refusal problems. When interviewers encounter a not-at-home situation, they are often tempted to substitute the chosen home with a neighboring dwelling where a respondent is available (or with the next number if a telephone survey is being conducted). This procedure may seem sound. However, it is only sound in terms of ensuring an adequate sample size. It may actually increase rather than decrease the nonresponse error. Consider the following example.

Example. Jane is an interviewer who is conducting personal interviews of women in order to ascertain their attitudes toward day-care centers for children. She has a list of households (as designated in the planned sample) that she is supposed to contact. It is 10 a.m. and Jane has already made contact with and interviewed three nonwork-

---

[7] For additional details about this study, see Leigh Lawton and A. Parasuraman, "The Impact of the Marketing Concept on New Product Planning," *Journal of Marketing*, 44 (Winter 1980), pp. 19–25.

Table 15.2 *Comparison of the Planned and Final Samples in the New-Product-Planning Study*

Industry Type	Composition of Planned Sample Number of Firms	 Percentage of Firms	Composition of Final Sample Number of Firms	 Percentage of Firms
Food and tobacco	46	9	11	10
Textile	43	9	12	11
Lumber and furniture	33	7	5	5
Paper and printing	43	9	8	8
Chemical and petroleum	38	8	7	7
Rubber, leather, and glass	40	8	9	8
Metal	94	19	15	14
Machinery	106	21	27	25
Transportation, measuring instruments, and miscellaneous	57	11	13	12
Total	500		107	

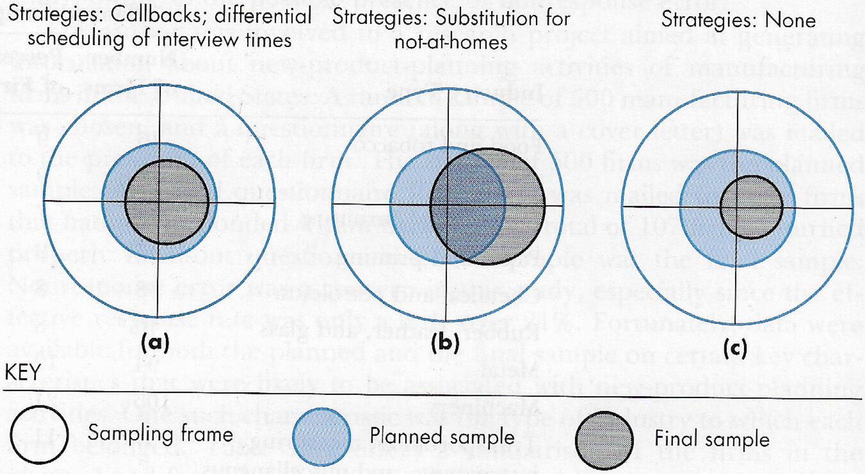
Note: All percentages have been rounded off; hence they may not add to exactly 100%.

ing women in the first three houses specified on her list. Jane is now at the fourth address on her list, but there is nobody home. However, she notices a similar house just across the street, and a couple of children are playing in the front yard of that house. The presence of children indicates somebody is home. Jane is wondering whether she should substitute the house across the street for the fourth house on her list. What do you think she should do?

If Jane does make the substitution, she will probably find another nonworking housewife, whereas the person who is not at home in the designated house is likely to be a working woman. If Jane makes a number of such substitutions (from the appearance of houses), the final sample will be biased because it will be dominated by nonworking housewives; the final sample, although similar in size to the planned sample, will be sharply different in terms of composition. The result will be serious nonresponse error. We can expect a woman's attitude toward day-care centers, which is the key variable in the study, to differ depending on whether or not she is a working woman.

The use of *callbacks* and *differential scheduling of interview times* to make contact with hard-to-reach respondents are more effective in reducing nonresponse error than the substitution of available respondents for the not-at-homes. For example, Belden Associates, a commercial marketing research firm, uses both these strategies to reduce nonresponse error in

Exhibit 15.4 *Effects of Strategies Used to Handle Not-at-Homes on Nonresponse Error*


Strategies: Callbacks; differential scheduling of interview times
Strategies: Substitution for not-at-homes
Strategies: None
(a)
(b)
(c)
KEY
Sampling frame
Planned sample
Final sample

personal-interview surveys that are a part of a Continuing Market Study (CMS) service offered to newspaper publishers. According to Belden Associates:

> In samples employing call-backs, interviewers must make a maximum of three calls (one original and two call-backs) to complete the interview with a selected respondent. . . . To reflect possible variations in consumer behavior by day of the week, interviewing is spread out Tuesday through Saturday (thus representing Monday-through-Friday "yesterday" newspaper exposure). To improve the chances of finding all types of people at home, interviewing is conducted approximately between 5:00 and 9:30 p.m. on weekdays and noon to 4:30 p.m. on Saturdays.[8]

Callbacks and differential scheduling (or spreading out) of interview times will add to the cost of a study. Moreover, they do not guarantee the completion of the total number of interviews specified in the planned sample. Nevertheless, nonresponse error in a study employing these techniques will be lower than in studies not employing them or in studies using substitution of respondents for not-at-homes, as illustrated by the diagrams in Exhibit 15.4. The amount of nonresponse error in Exhibit 15.4(a) is lower than it is in Figure 15.4(b) or 15.4(c). Also notice that the nonresponse error in Exhibit 15.4(b) is higher than in Exhibit

---

[8] "The Belden Continuing Market Study: The Indianapolis '74 Newspaper Audience," a brochure prepared by Belden Associates, 2900 Turtle Creek Plaza, Dallas, TX 75219, p. D in the "Technical Notes" section.

15.4(c), despite the final sample size in 15.4(b) being as large as the planned sample size.

The respondent-refusal problem is difficult to avoid, especially in telephone surveys. At least in personal-interview surveys, once the interviewer comes face to face with a respondent, it may be psychologically hard for the respondent to say no. In telephone surveys, though, the respondent can easily decline to participate by just hanging up. There are two reasons for respondent refusals: (1) The time when the interviewer makes contact is inconvenient for the respondent, and (2) the respondent is reluctant to participate for reasons other than convenience.

Refusals stemming from inconvenient interview timing can be handled to some extent in a fashion similar to the handling of not-at-homes—through callbacks and judicious scheduling of the initial contacts at times likely to be convenient to respondents. Refusals occurring because of respondent resistance are tougher to handle. However, offering incentives (monetary or nonmonetary) and using well-qualified interviewers who are trained to overcome respondent objections may be helpful.[9]

Refusals due to respondents' resistance to the survey itself have a special bearing on the plus-one and random-digit dialing strategies for overcoming sampling frame errors in telephone surveys. While telephone surveys using these strategies may be able to minimize sampling frame error, research evidence (and perhaps even intuition) indicates that these surveys may still encounter serious nonresponse error owing to refusals. For instance, one of the conclusions of the research project conducted by Pacific Northwest Bell that we mentioned earlier was as follows:

> Even if respondents with nonpub[lished] numbers are contacted, they are more likely to refuse to be interviewed, and less likely to answer all the questions than are respondents with listed telephone numbers. Evidently, their desire for privacy extends to surveys. These response rate data provide clues to the characteristics of nonrespondents. Nonrespondents to surveys probably differ from respondents in systematic ways.[10]

Nonresponse error can be minimized, but it cannot be completely avoided in most surveys. Sound research practice requires careful consideration of this potential limitation in drawing inferences and making recommendations. Moreover, a good research report must explicitly disclose the implications of nonresponse error (or any other error, for that matter). Disclosure will enable the report's audience to exercise caution in interpreting the research findings.

---

[9] For example, one technique that has been used with some success to overcome respondent objections is the foot-in-the-door technique, see Jonathan L. Freedman and Scott C. Frazer, "Compliance Without Pressure: The Foot-in-the-Door Technique," *Journal of Personality and Social Psychology,* October 1966, pp. 195–202.

[10] Moberg, "Biases in Unlisted Phone Numbers," pp. 54–55.

Incorporating Effects of Nonresponse Errors into Sample Results. Our discussion thus far focused on procedures for minimizing nonresponse errors and for ascertaining whether such errors could be a problem in studies just completed. In some situations the researcher may be able to estimate the effects of nonresponse errors on at least certain key variables and then adjust the sample results accordingly.[11]

For instance, suppose only n sample units from a planned sample of N units respond to a survey dealing with recreation. Let the mean value for a key variable (e.g., expenditures per year on record albums) be $\overline{X}_1$ for respondents and $\overline{X}_2$ for nonrespondents—and, of course, at this stage $\overline{X}_2$ is unknown. The correct mean value for the planned sample (say $\overline{X}$) is

$$\overline{X} = \frac{n\overline{X}_1 + (N - n)\overline{X}_2}{N}.$$

Sometimes, a rough estimate of $\overline{X}_2$ may be available through secondary sources. For instance, when some information about the demographic profile of nonrespondents is available, past studies or music industry sources can be consulted to obtain an approximate estimate of a variable such as expenditures on record albums made by the nonrespondent group. If $\overline{X}_2$ differs greatly from $\overline{X}_1$, the formula can be used to adjust $\overline{X}_1$ to obtain a more accurate estimate of the sample mean $\overline{X}$.

Alternatively, an attempt can be made to estimate $\overline{X}_2$ through an intensive follow-up of a small random sample of nonrespondents. For instance, a sample of nonrespondents to a mail survey can be contacted by telephone or in person and requested to answer a relatively small set of key questions on the original survey. Data thus gathered can be used to estimate $\overline{X}_2$. Again, if $\overline{X}_2$ differs greatly from $\overline{X}_1$, the formula can be used to adjust the results obtained from the initial group of respondents. The ability to make such an adjustment depends on factors such as whether nonrespondents can be identified, whether adequate resources are available for intensive follow-up efforts, and whether nonrespondents cooperate when recontacted. In any case, the preceding discussion highlights an important point. When a researcher does not make any

---

[11] For references related to statistical treatment of the effects of nonresponse errors, see *Statistical Adjustment for Nonresponse in Sample Surveys: A Selected Bibliography with Annotations* (Monticello, Ill.: Vance Bibliographies, 1979); see also James C. Ward, Bertram Russick, and William Rudelius, "A Test of Reducing Callbacks and Not-at-Home Bias in Personal Interviews by Weighting At-Home Respondents," *Journal of Marketing Research*, 22 (February 1985), pp. 66–73.

Research in Use 15.2 *Channelview Cable Company*

Adjusting Results for Nonresponse Errors

Channelview Cable Company (CCC) located in a large metropolitan area in the Southwest, conducted a mail survey of 1000 households randomly chosen from its viewing area. The response rate for the survey was 30% (i.e., 300 households returned completed questionnaires). While the survey covered a wide variety of topics, two pieces of information that CCC was particularly interested in were (1) the number of hours per week that a typical household spent watching television, and (2) the proportion of households owning one or more VCRs. Analysis of data obtained from the 300 respondents revealed the following: (1) average number of hours of TV viewing per week per household = 21 hours, and (2) proportion of households with VCRs = .6.

Since 70% of the households that received the survey questionnaire did not respond, CCC was concerned about the accuracy of the aforementioned results. It therefore commissioned a brief telephone survey of 100 nonresponding households to collect data on just the two critical issues. Data from this telephone survey revealed that the average number of hours of TV viewing per week per household was 14 and that 40 of the households had VCRs (i.e., the proportion of households with VCRs was .4). From these data, the findings from the mail survey were adjusted as follows to yield more accurate estimates:

$$\text{Hours of TV viewing per week per household} = \frac{(300)(21) + (700)(14)}{1000} = (.3)(21) + (.7)(14) = 16.1,$$

$$\text{proportion of households with VCRS} = \frac{(300)(.6) + (700)(.4)}{1000} = (.3)(.6) + (.7)(.4) = .46.$$

adjustment of $\overline{X}_1$, a critical assumption being made is that $\overline{X}_1$ is not significantly different from $\overline{X}_2$. The accuracy of the sample results and the inferences stemming from them will be questionable when this assumption is not valid.

Research in Use 15.2 illustrates the correction of mail survey estimates by using data from a telephone follow-up of a sample of nonrespondents. Two points that are implied in this illustration are noteworthy. First, the correction formula can be used to adjust sample means as well as proportions. Second, the follow-up procedure must generate data on each variable or question that needs to be "adjusted."

Data Error

Data errors due to respondents' providing distorted answers (either intentionally or unintentionally) are difficult to control directly. However, an examination of why respondents distort their answers may reveal strategies for minimizing such errors. The motivation for deliberately distorting responses is likely to be high when the respondent perceives a question to be sensitive or a question invites a socially desirable answer; the respondent is irritated by the questionnaire or the questioning process as a whole and wants to frustrate the researcher; and/or a question is difficult for the respondent to answer. How can one avoid, or at least minimize, the incidence of these conditions?

Questions about sensitive issues obviously cannot be avoided if the purpose of a research project requires that such issues be examined. However, the *degree* of sensitivity of the questions as perceived by the respondents can be minimized by carefully adhering to the guidelines for good questionnaire construction, which we discussed in Chapter 10. For instance, of the following two alternative wordings of an income question, a respondent will perceive alternative A to be somewhat more sensitive than alternative B.

A. What is your annual gross income (before taxes)?

$________________

B. Indicate your annual gross income (before taxes) by checking the appropriate category:

________ Less than $10,000

________ $10,000 to less than $15,000

________ $15,000 to less than $20,000

________ $20,000 to less than $25,000

________ $25,000 to less than $30,000

________ $30,000 to less than $35,000

________ $35,000 to less than $40,000

________ $40,000 or more

Question B will only yield categorical data (i.e., nonmetric data) on income; question A should ideally yield more refined income data (i.e., metric data). However, since question A is somewhat more threatening than question B from the respondent's point of view, the answers to it may be seriously distorted. Nevertheless, the apparent refined nature of the data may give the researcher a false sense of accuracy and prompt the computation of statistics such as the mean income. However, if the data are distorted, the computed statistics will be biased, resulting in erroneous inferences. Another potential drawback of question A is that more respondents are likely to refuse to answer it than to answer question B.

A variety of other strategies, several of which were described under questionnaire design (Chapter 10), are available to minimize deliberate as well as unintentional response distortions. These strategies include the following techniques.

1. Train interviewers to establish rapport with respondents.
2. Choose interview times that are convenient for respondents.
3. Avoid the use of leading questions (i.e., questions that invite a certain type of response irrespective of how a respondent actually feels).
4. Do not ask sensitive questions early in the interview.
5. Make the questionnaire interesting and easy for respondents.
6. Pretest the questionnaire.

Taking precautions to minimize data errors caused by respondents is especially important, since such errors are much harder to detect than sampling frame or nonresponse errors. As we established earlier, for sampling frame and nonresponse errors one can indirectly ascertain their extent after a study has been completed—for example, by comparing the sampling frame with the ideal population and/or by comparing the planned sample with the final sample—and modify one's inferences accordingly. No such after-the-fact adjustment of inferences is possible when there are data errors caused by respondents because of the difficulty in uncovering the presence and the magnitude of such errors.

A potential source of error in telephone and personal-interview surveys, as well as in observation studies, is the interviewer. (Although we will be using the term *interviewer* hereafter, some of the interviewer errors we will discuss may also be relevant for studies using the observation method to gather data.) The keys to minimizing data errors stemming directly from interviewers are *selection*, *training*, and *supervision*. Substantial data errors may result if enough time and energy are not devoted to these activities. Thus serious efforts must be made *to prevent* (rather than *to cure*, after the fact) interviewer errors. Similar to the errors made by respondents, data errors caused by interviewers are hard to detect and measure. One possible exception is *interviewer cheating*, which can be detected in some situations, as we will see shortly. However, even if interviewer cheating is detected, the researcher usually only has the option of discarding (rather than correcting) the data generated by the interviewer. While this option may be better than inadvertently using the erroneous data, it is also an expensive option owing to data wastage.

Recognizing the seriousness of interviewer biases, many reputable marketing research firms take elaborate steps for controlling such biases. For example, according to Walker Research, Inc.:

> We believe comprehensive training is a key factor in data collection quality and have developed training programs that are among the best in the industry. Walker training programs include a unique three-level interviewer training, specialized study training, CRT [Cathode Ray Tube] interviewing training, supervisor training and travelling supervisor training.
>
> Interviewer candidates throughout our network are evaluated for verbatim reading and listening skills. Only those candidates who meet Walker's rigid standards are hired and trained to probe, clarify, and record responses the WALKER WAY.[12]

As another example, Louis Harris and Associates offers it clients a regular service called Harris Perspective Service, which is designed

> to provide corporations with timely and relevant information necessary to evaluate the social and political climate they operate within, and to help make sound business decisions regarding corporate policy, planning, marketing, and communications.[13]

This service is based on data obtained through personal interviews of a random sample of the American public. Before using the collected data, the firm takes specific steps to detect and handle interviewer cheating:

> Interviews are validated for each Perspective study. Interviewers send the name, address, and phone number (if available) of respondents to a validation service selected by the Harris firm. In New York, the Harris firm makes up a list of five study questions which are given to the validation service. Approximately 20% of each interviewer's respondents, randomly chosen, are contacted by the validation service (generally by telephone, but by mail if necessary) and asked the study questions. If there is any doubt about the authenticity of an interviewer's work, all of his or her interviews are carefully checked.[14]

In addition to proper selection, training, and supervision, the following techniques can also be helpful in reducing sloppiness on the part of interviewers and the incidence of interviewer cheating:

- Keep the interviewers' tasks as simple as possible and make sure they clearly understand it. For instance, a complex questionnaire or a complicated sample selection procedure can be very frustrating to interviewers and may motivate them to cheat. As Boyd, Westfall, and Stasch correctly observed: "Many investigators are coming to the conclusion that cheating is as much a problem of morale as of morals."[15]

---

[12] "The Walker Way . . . to Quality Research," a brochure prepared by Walker Research, Inc., 3939 Priority Way South Drive, Indianapolis, IN 46280-0432.

[13] "Opinion Research for Business: Harris Perspective Service 1979," a brochure prepared by Louis Harris and Associates, 630 Fifth Avenue, New York, NY 10020, p. 28.

[14] Ibid.

[15] Harper W. Boyd, Ralph Westfall, and Stanley F. Stasch, *Marketing Research Text and Cases* (Homewood, Ill.: Irwin, 1985), pp. 447–448.

- Provide interviewers with adequate compensation. Contacting strangers, convincing them to participate, and eliciting responses from them is not easy, especially in a face-to-face situation and especially when a study deals with sensitive subject matter. The mental and physical strains on interviewers are likely to be quite high. Consequently, if they think their compensation is inadequate, the chances of their cutting corners during the fieldwork are likely to increase.

Some obvious data errors, such as inconsistent responses to different questions, can be detected through the use of appropriate editing and coding procedures (more will be said on this subject in the next chapter). For instance, in response to a question about store patronage, if a respondent claimed to have never been to store ABC, but later in the questionnaire acknowledged buying brand X, which is only available in store ABC, there is obviously data error in the form of response distortion. Such errors have to be eliminated before data analysis can begin. The best way to avoid errors in analyzing the data and interpreting the findings is to understand the nature of the data (e.g., whether they are ordinal or interval) and the quality of the data (e.g., whether they suffer from any nonresponse error) *before* one begins the data analysis. Researcher objectivity is crucial in minimizing errors in the data analysis and interpretation stages of a research project.

SUMMARY

The total error in a research project consists of sampling and nonsampling errors. Table 15.3 recapitulates the different types of errors and summarizes strategies for controlling them.

Sampling error is random and is quantifiable if a probability-sampling method is used. Moreover, sampling error can be reduced by increasing the sample size and/or by using a rigorous sampling procedure with high statistical efficiency.

Nonsampling errors introduce systematic biases that are usually quite hard to detect, let alone measure. Three major forms of nonsampling errors are sampling frame error, nonresponse error, and data error. These errors stem from a variety of sources. The popular saying that prevention is better than a cure is particularly pertinent to handling many forms of nonsampling error that are difficult to uncover and correct after they have occurred.

A final point worth emphasizing is that the numerous errors that can occur in a research project have implications not only for researchers but also for research users (managers or decision makers in general). Very often decision makers are content with knowing that a study employed random sampling and surveyed a large sample of units. However, as we have seen, these criteria are not capable of capturing the total

Table 15.3 *Summary of Potential Errors in Research Projects*

Type of Error	Ways of Minimizing Error
Sampling Error (Occurs when a sample, rather than the entire population, is studied)	Increasing the sample size Using a statistically efficient sampling plan; i.e., making the sample as representative of the population as possible
Nonsampling Error	
Sampling frame error (Occurs when the sampling frame is not representative of the ideal population)	Starting with a complete sampling frame Modifying the sampling frame to make it representative of the ideal population (e.g., using plus-one dialing in telephone surveys)
Nonresponse error (Occurs when the final sample is not representative of the planned sample)	Mail surveys: Increasing response rates through the use of incentives, follow-up mailings, etc. (*Caution*: Increase in response rate per se may not reduce nonresponse error) Telephone and personal-interview surveys: Making callbacks and spreading out the time blocks during which interviews are conducted
Data error (Occurs because of distortions in the data collected as well as mistakes in data coding, analysis, or interpretation)	Ensuring that the questionnaire is good—i.e., is simple, unbiased, etc. Proper selection, training, and supervision of interviewers Keeping the interviewers' task simple and making it clear to them Giving adequate compensation to the interviewers Using sound editing and coding procedures Taking into account the nature and quality of the data in analyzing and interpreting them

error that can occur in a research project. Hence a decision maker should carefully examine a number of issues, such as the following questions, before drawing any conclusions from the research findings.

- What sampling frame was used?
- What are the characteristics of the respondents?
- What follow-up methods were used to reach nonrespondents?
- How do the respondents compare with nonrespondents on characteristics relevant to the research study?
- What were the extent and the nature of the training and supervision received by the interviewers?

A good research report should already have the answers to questions like these. If a report fails to address methodological issues adequately, a decision maker must ask the researcher(s) for additional information. Such information will be very helpful in determining how much faith one can have in the research findings. Similarly, even when evaluating a research proposal, decision makers will benefit from asking probing questions about whether and to what extent controls have been built into the proposed research design in order to handle potential errors.

KEY TERMS

Sampling error
Nonsampling error
Total error
Unbiased estimate
Sampling frame error
Nonresponse error
Not-at-home problem
Respondent-refusal problem
Data error
Plus-one dialing
Random-digit dialing

QUESTIONS

1. "The higher the statistical efficiency of a sampling procedure, the higher will be the quality of the data collected from a sample chosen by using that procedure." Discuss this statement.
2. Define *sampling error* and *nonsampling error*. What effect will these errors have on the estimate of a population parameter value from sample estimate values?
3. Define and briefly discuss *sampling frame error*. Give an example of your own to illustrate it.
4. Define *nonresponse error*. Will nonresponse error occur whenever certain sample units fail to respond to a survey? Why or why not?
5. State and briefly describe the two main sources of nonresponse error in personal and telephone surveys.
6. Define *data error*, and illustrate it with two different examples.
7. Briefly discuss the ways in which sampling frame error can be reduced. Illustrate your answer with examples.

8. Name two ways in which the response rates to mail surveys can be increased. Does the resulting higher response rate mean lower nonresponse error? Why or why not?

9. In what ways can the not-at-home problem be handled? Which of these ways would you recommend? Why?

10. "Nonresponse error can be reduced but not eliminated in most surveys." Discuss this statement.

11. A questionnaire pertaining to home appliances was mailed to a random sample of 2000 households. Completed questionnaires were received from 500 households of which 200 indicated that they had microwave ovens. A telephone follow-up of a random sample of 150 nonresponding households indicated that 30 of them had microwave ovens. On the basis of these data, what is the best estimate of the proportion of households having microwave ovens?

12. Discuss the role that good questionnaire design can play in reducing nonsampling errors.

13. Develop an argument in support of the following statement: "Of the three major forms of nonsampling error, data error is the most serious."

14. Briefly describe the various interviewer-related strategies that can be used to reduce nonsampling errors.

15. Can the research user play any role in ensuring that the information generated from a research project is of high quality? Explain your answer.

Cases for Part Five

CASE 5.1 *DEXTER'S DEPARTMENT STORES*

CASE 5.2 *NORTHWEST BANK OF WOODCREEK (A)*

CASE 5.3 *PHOENIX NEWSPAPERS, INC.*

CASE 5.4 *MCCAW CABLEVISION*

CASE 5.1

Dexter's Department Stores

Dexter's Department Stores is a chain of 150 upscale department stores operating in ten different states in the United States. The company's president and CEO, Don Dexter, has been concerned for quite some time about a growing phenomenon in the field of retailing in general and in Dexter's case in particular—namely, that fewer male business graduates are attracted to careers in retailing than female business graduates. While over half of all business graduates are males, apparently less than 20% of them are considering retailing as an attractive career option.

THE COMPANY'S RECRUITING EFFORTS

Every year, Dexter's Department Stores recruits graduates from several major universities in each of the ten states in which it has operations as well as from a few well-known universities outside these states. In all, Dexter's recruiters visit 40 college campuses regularly—at least one each year. They also recruit at a few other campuses but not on a regular basis. During the past year they filled 50 job openings at Dexter's Department Stores with business graduates. All 50 graduates were from the 40 campuses the recruiters visited regularly. Of these 50 graduates only 8 were males.

RESEARCH PROBLEM

Mr. Dexter has been quite pleased with the quality of the female graduates his company hired. In fact, several of the females hired within the past three or four years have rapidly risen to high-level managerial positions as a result of their excellent performance. Nevertheless, Mr. Dexter was puzzled and intrigued by the less-than-enthusiastic response from male business school graduates to his company's supposedly aggressive recruiting efforts on college campuses. He wanted to know why a highly successful company like Dexter's Department Stores was having difficulty in attracting male business graduates. Specifically, he wanted to gain insight into the relative perceptions of male and female business school seniors about a career in retailing in general and one with Dexter's Department Stores in particular. He also wanted to get a feeling for how male and female business school seniors rated retailing careers vis-à-vis careers in such fields as selling and advertising.

Mr. Dexter therefore asked his director of personnel to look into the broad issue of female versus male college seniors' views about retailing careers. The director of personnel, in turn, requested a commercial marketing research firm to draw up a research proposal for examining the issues that Mr. Dexter was interested in. Several weeks later, the research firm submitted a proposal that detailed a formal, structured telephone survey of 1000 business school seniors across the country. The procedure for selecting the 1000 respondents suggested by the research firm is described in the next section.

SAMPLING PLAN

The section of the proposal that dealt with sampling read as follows:

> We will first obtain a list of all universities/colleges in the United States that have a four-year degree program in business administration. We expect that there will be about a thousand such universities. We will pick a systematic sample of 20 universities from this list. If there are 1000 universities in the list, for example, we will pick every 50th university on the list, after randomly picking the first one.
>
> We will then get the student telephone directories from the 20 chosen universities. From each telephone directory five pages will be chosen at random, and from each chosen page 10 business school seniors' names and telephone numbers will be chosen at random for conducting the survey. Appropriate adjustments will be made in this last stage to ensure that 5 of the 10 seniors selected from each page are males.
>
> In summary, 50 business school seniors—25 males and 25 females—will be randomly chosen from each of 20 college campuses, which, in turn, will have been chosen from a list of all campuses with four-year business programs. The resulting total sample of 1000 seniors will give us an adequate, unbiased cross section of students for the telephone survey.

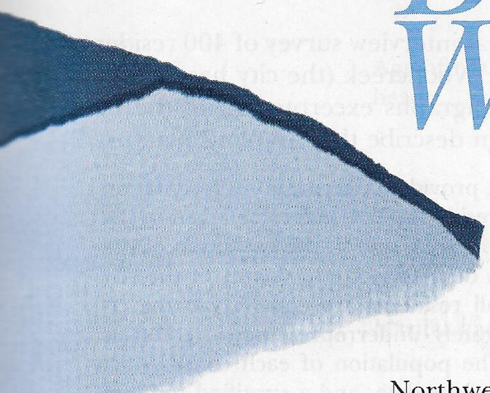
QUESTIONS

1. How would you describe the proposed sampling plan in terms of the various sampling techniques we have presented? Explain your answer.

2. Critically evaluate the proposed sampling plan. What are its advantages and disadvantages? Is the plan appropriate for or consistent with the types of information Mr. Dexter desires?
3. What alternative sampling plan would you propose? Why?
4. Comment on the appropriateness of conducting a structured telephone survey of 1000 seniors. If you think it is appropriate, explain why. If you think it is inappropriate, what alternative approach would you suggest, and why?

CASE 5.2

Northwest Bank of Woodcreek (A)



Northwest Bank of Woodcreek (NBW) is the largest of three banks in Woodcreek, a city with a population of 120,000. The NBW recently asked a local marketing research firm to conduct a competitive study of area financial institutions. Specifically, the top management of NBW wanted information on the financial-patronage habits of local residents and their perceptions of financial institutions in Woodcreek. To satisfy top management's overall information needs, the research firm specified the following more detailed objectives:

1. The investigate the market shares of the financial institutions in Woodcreek for several types of financial services.
2. To explain why Woodcreek residents chose to deal with the financial institutions they are presently associated with.
3. To determine the extent of bank switching and the reasons for switching.
4. To uncover the perceived advantages and disadvantages of the three primary institutions in Woodcreek, banks, savings and loan associations, and credit unions.
5. To determine the extent of awareness and use of, and interest in electronic-banking services.
6. To examine the extent of recognition for the slogans and the symbols of the banks and the savings and loan associations in Woodcreek.
7. To determine the perceived images of the major Woodcreek banks.
8. To examine the demographic profiles of customers of the various financial institutions.

9. To describe the patronage patterns in Woodcreek, census tract by census tract.
10. To determine if and how the customers of rival institutions can be attracted to Northwest Bank of Woodcreek.

SAMPLING PLAN

The research firm proposed a personal-interview survey of 400 residents chosen from various census tracts in Woodcreek (the city has a total of 25 census tracts). The following paragraphs excerpted from the final report submitted by the research firm describe the sampling plan used.

> So that the results of this study might provide an accurate representation of the true viewpoints of Woodcreek residents, care had to be taken in the questionnaire design, in the sample selection, and in the execution of the study. The population of interest was defined as the residents of the city of Woodcreek. However, since not all residents are equally valuable as banking customers, the study deliberately underrepresented the census tracts that were low-income areas. The population of each census tract was determined from United States census data, and a stratified random sample was drawn from the Woodcreek City Directory.
>
> Our sample was drawn so as to have a composition similar to that of the population of Woodcreek, with the exception of the planned underrepresentation of certain low-income census tracts. Census tracts 1 and 23 were eliminated because of their low-income level and very small population. In the remaining low-income tracts (2, 3, 6, 7, 9, and 18) half as many respondents were drawn as would have been had these tracts been fully represented in the sample [a map of Woodcreek showing the various census tracts was included in the report].
>
> Professional interviewers were used to conduct the personal interviews.

ACTUAL SAMPLE

The following description of the sample that provided the data for the study appeared in the research firm's final report.

> A total of 393 usable questionnaires were obtained from the 400 residents surveyed. Overall, the demographic characteristics of the sample surveyed seemed to have met our expectations. Consequently, we feel confident that the sample is representative enough to draw meaningful conclusions from the information obtained. As can be seen from an examination of the demographic characteristics contained in Table 1, all demographic groups are adequately represented. The 65-and-over age-group is slightly overrepresented, and the under-25 age-group is somewhat underrepresented. Feedback from our interviewers indicated that this result occurred because of the differences in the availability and willingness of these age-groups to participate in the survey. However, this slight deviation is not serious enough to significantly affect the conclusions drawn from the study.

Table 1 *Demographic Characteristics of the Sample Surveyed*

Characteristics	**Categories**	**Percentage of Respondents in Each Category**
Sex	Male	36.6
	Female	63.4
Age of household head	Under 25	7.9
	25–34	21.1
	35–44	19.6
	45–54	16.3
	55–64	16.3
	65 and over	17.6
	Refused to answer	1.2
Marital status	Married	79.9
	Single (includes separated or divorced, and widowed)	20.1
Education level of highest-educated spouse	Less than high school	6.9
	Attended high school	8.7
	High school graduate	44.0
	High school + some college	20.1
	College degree or beyond	19.6
	Refused to answer	0.7
Household annual income	Less than $5,000	16.0
	$5,001–$10,000	15.8
	$10,001–$15,000	21.1
	$15,001–$20,000	19.1
	$20,001–$25,000	7.4
	$25,001–$30,000	5.1
	Over $30,000	3.3
	Refused to answer	12.2
Length of stay in Woodcreek	Less than 6 months	3.1
	6 months–less than 1 year	1.0
	1 year–less than 2 years	2.3
	2 years and over	93.6

Note: All numbers in this table represent percentages based on a total sample of 393 respondents.

QUESTIONS

1. What label would you place on the sampling plan used by the research firm? Critically evaluate the sampling plan in light of the objectives of the study and the composition of the actual sample obtained.
2. If you were to do this study over again, what specific sampling procedure(s) would you use? Justify your answer.

CASE 5.3

Phoenix Newspapers, Inc.

Phoenix Newspapers, Inc., publisher of two major Phoenix area newspapers—the *Arizona Republic* and the *Phoenix Gazette*—has a long tradition of community involvement and service. As a service to local organizations and the business community, the company publishes an annual brochure called "Inside Phoenix." This brochure is based on a comprehensive study of Phoenix area residents and contains a wealth of information particularly useful as a resource tool for planning and marketing. The annual Phoenix Market Study on which "Inside Phoenix" is based is conducted by the Marketing Research Department of Phoenix Newspapers, Inc. The year 1982 marked the 28th publication of "Inside Phoenix." A description of the methodology used in the 1982 Phoenix Market Study is excerpted here from "Inside Phoenix 1982."

PHOENIX MARKET STUDY METHODOLOGY

The 1982 Phoenix Market Study is a comprehensive media/market study of the Phoenix SMSA (Standard Metropolitan Statistical Area). A broad range of data including demographic characteristics, newspaper, radio and television audiences, and product and shopping data were collected from a common group of respondents. The wealth of available data provides the opportunity for a wide range of analyses, including media penetration singularly and in schedules, market segmentation, product and audience profiles, cross-shopping patterns, and numerous other cross-tabulations of the data.

In order to obtain this data, three different questionnaires were used.

Case material was provided courtesy of Phoenix Newspapers, Inc., P.O. Box 1950, Phoenix, AZ 85001.

An initial telephone interview was used to collect audience data on radio, television, daily and weekly newspapers and magazines. In addition, travel data and grocery shopping information along with demographic characteristics were obtained.

Each respondent was mailed a self-completion questionnaire which contained questions about leisure time and vacation activity, stores and shopping centers, automotive equipment, alcoholic beverage and other market products areas.

Finally, each respondent was telephoned a second time. This second telephone interview was, first of all, a validation of the initial telephone interview. In addition, the daily and Sunday newspaper questions were asked again, thus producing a large data base for audience estimates, and also the necessary input for estimating multiple reach and frequency.

SAMPLE DESIGN AND EXECUTION

The survey represented all adults 18 years of age and older residing in telephone households in Maricopa County. Persons living in nonhousehold types of accommodations, such as student dormitories, military bases, and institutions of various types, were excluded.

The interviewing was carried out between September 9 and October 17, 1981. All interviewing was done from Tuesday through Saturday, primarily in the late afternoon and evening to accommodate working men and women.

Random-digit dialing was used to represent telephone households. A random selection of four-digit numbers was selected from each exchange, proportionate to the household listings in that exchange. These were used as "seed" numbers, to which random numbers were added. This produced a random sample of both listed and unlisted numbers. Within each telephone household one adult was randomly designated as the respondent. No substitution was permitted. Callback procedures at different times of the day and different days of the week were used to obtain interviews with designated respondents. The completed sample representing 61.3% of the sample frame includes 2313 respondents, of whom 1152 are men and 1161 are women.

A self-executing questionnaire concerning shopping and purchasing behavior was mailed to each eligible respondent. A dollar bill was used as an incentive and a reminder mailing was used as a prompt. 1518 questionnaires were obtained using these methods, or 65.6%.

Each respondent was telephoned a second time at least eight days after the initial interview. The second interview was a validation technique and also provided additional newspaper audience data completed with 1869 respondents, or a completion rate of 80.8%.

SAMPLING DEVIATIONS

It should be remembered that, as with all surveys, the findings are subject to the ever-present laws of chance. As such, there will be a few findings that will depart from the true levels by an uncomfortably wide margin.

Fortunately, by these same laws of chance, the number of these departures will be relatively small. Users of the data should bear in mind that the findings are subject to sampling tolerance, and these tolerances should be taken into consideration when critical evaluations are made.[1]

QUESTIONS

1. Critically evaluate the sampling procedure used in conducting the 1982 Phoenix Market Study.
2. Quantify the sampling tolerance (i.e., sampling error) associated with percentage estimates derived from data provided by (a) all respondents; (b) male respondents; (c) female respondents.
3. Are there likely to be any errors other than sampling error in the 1982 Phoenix Market Study? If so, what are they, and how serious are they likely to be?

---

[1] Reprinted with permission from *Inside Phoenix 1982* published by Phoenix Newspapers, Inc.

CASE 5.4

McCaw Cablevision

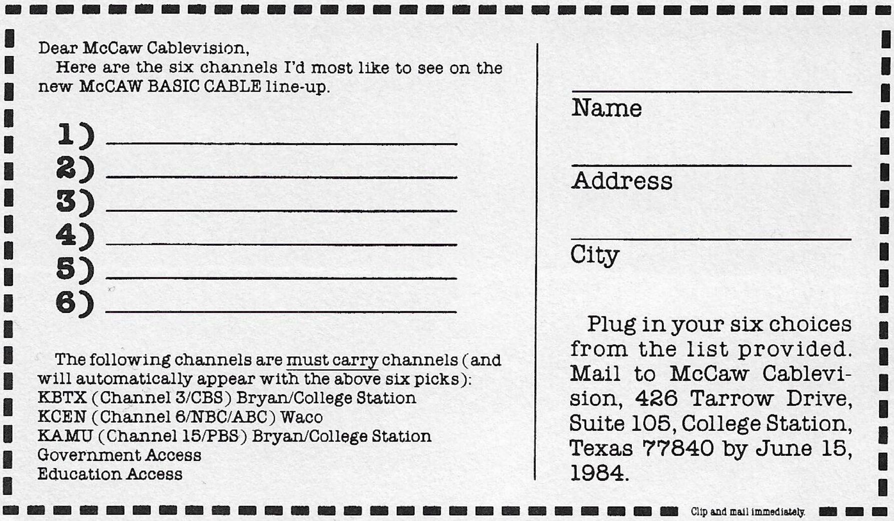
The community of Bryan–College Station, located in central Texas, was served for a number of years by two competing cable television companies, Community Cablevision and Midwest Video Corporation. But in early 1984 both companies were bought by McCaw Communications, Inc., a company based in Bellevue, Washington. McCaw Communications decided to merge the two companies into a single unit called McCaw Cablevision.

One of several potential problems associated with merging the two independent companies was that the companies had been offering their respective subscribers somewhat different sets of channels as part of their basic channel packages. McCaw Cablevision therefore had to decide which channels should be included in, and which should be excluded from, a new basic package to be offered to all subscribers. To aid in designing its basic package, McCaw Cablevision decided to seek input from the community in the following way.

In late May 1984 McCaw Cablevision bought a full-page ad in the Bryan–College Station *Eagle,* the community's daily newspaper. This ad briefly described the merger and sought the public's preferences concerning a number of different channels that could be included in the basic package. Specifically, part of the ad read as follows:

> McCaw is merging the two previous cable company line-ups into a single lineup. And, after counting in the channels we're required to carry by federal and local rules, we have room left for six others. So we want your favorites.
>
> We can't guarantee each of your choices will make our final list. But we'll take the six most popular picks, then plug them into our new McCAW

Exhibit 1 *McCaw Cablevision Coupon*



Dear McCaw Cablevision,
Here are the six channels I'd most like to see on the new McCAW BASIC CABLE line-up.

1) ______
2) ______
3) ______
4) ______
5) ______
6) ______

The following channels are must carry channels (and will automatically appear with the above six picks):
KBTX (Channel 3/CBS) Bryan/College Station
KCEN (Channel 6/NBC/ABC) Waco
KAMU (Channel 15/PBS) Bryan/College Station
Government Access
Education Access

Name ______
Address ______
City ______

Plug in your six choices from the list provided. Mail to McCaw Cablevision, 426 Tarrow Drive, Suite 105, College Station, Texas 77840 by June 15, 1984.

Clip and mail immediately.

Source: Courtesy of McCaw Communications Companies, Inc., Bellevue, Wash.

> BASIC CABLE line-up. And soon you'll be watching your favorite channels on your new McCAW BASIC CABLE.[1]

The text of the ad was accompanied by a list of 19 different cable channels. Readers were requested to select 6 of these channels they would most like to see on the new McCAW Basic Cable lineup and to list those preferences on a coupon that accompanied the ad. A copy of the coupon is shown in Exhibit 1.

QUESTIONS

1. What sampling frame and sampling method are implied in the procedure used by McCaw Cablevision to seek the community's cable channel preferences?
2. What is your evaluation of the procedure used and the usefulness of the data generated through it? Justify your answer. If you feel the procedure was not sound, what do you think should have been done differently?

---

[1] Bryan–College Station *Eagle,* May 27, 1984, p. 3A.

PART SIX

Data Analysis

CHAPTER 16 *Quality Control and Initial Analysis of Data*

CHAPTER 17 *Hypothesis Testing and Its Role in Data Analysis*

CHAPTER 18 *Techniques for Examining Associations: Chi-square, Correlation, and Regression*

CHAPTER 19 *Overview of Other Multivariate Techniques*

CASES FOR PART SIX

CHAPTER 16

Quality Control & Initial Analysis of Data

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Define editing and distinguish between field edit and office edit.*
- *Define coding and outline the steps involved in it.*
- *Identify the different types of "don't know" responses and discuss the ways in which they can be handled in the coding process.*
- *Define case, data record, and data set.*
- *Distinguish between a raw variable and a transformed variable.*
- *State the potential uses of one-way tables.*
- *Describe what a skewed distribution is and illustrate it with an example.*
- *Construct and interpret two-way tables.*
- *Indicate the precautions to be taken in interpreting two-way tables.*
- *State the potential difficulties in constructing and interpreting multiway tables.*

As demonstrated in the previous chapter, a variety of errors can occur during fieldwork. While certain fieldwork errors can be prevented through precautionary measures such as proper questionnaire design and adequate interviewer training, unanticipated errors can and frequently do occur. Failure to quickly discover and remedy those errors will result in meaningless or misleading data.

For example, a descriptive-research project the author was associated with involved a personal-interview survey of a quota sample of 500 adult respondents. The purpose of the project was to obtain data on life-style variables like leisure time activities, reading habits, and television-viewing behavior. Several seemingly professional interviewers were recruited to gather the data. Sex was one of the quota controls, and the interviewers were instructed to divide their interviews evenly between males and females. The questionnaire, though lengthy, was fairly structured and straightforward. Moreover, no difficulties or complaints were voiced by the interviewers during data collection. As a result, the researchers in charge grew complacent and did very little checking of the completed questionnaires immediately after receiving them from the field. The only periodic checking they did was a cursory examination of interviewers' quota sheets to verify whether adequate progress was being made toward filling the various cell quotas. The questionnaires were not scrutinized until well after all of them had been returned.

A careful examination of the questionnaire data revealed that 25 questionnaires had *both* male and female categories checked under the item "respondent's sex." Yet the interviewers had been assigned quotas of 50% male and 50% female respondents, and these quotas had apparently been adhered to. Additional investigation traced all 25 questionnaires to the same interviewer, who, unfortunately, had moved out of the area by that time. Therefore the only clue to these strange responses was the following additional notation the interviewer had made on one of the questionnaires:

__X__ Male Husband; answered most of the questions

__X__ Female Wife; answered certain leisure activity questions

Evidently, this interviewer did not restrict the interview to just one person in households where both husband and wife were willing to participate in the survey. Conducting part of the interview with a male and part of it with a female was a serious error; all 25 interviews had to be discarded. However, this mistake might have been an honest mistake since the formal quota-sampling instructions asked the interviewers to *contact* (apparently interpreted by this particular interviewer as "make initial contact with") 50% males and 50% females. The point is, irrespective of whether the mistake was an honest one, it could, and *should,* have been caught and corrected much earlier than it was. Discovering fieldwork errors and attempting to correct them early in the project is discussed in the next section.

EDITING

Definition. ***Editing*** is the process of examining completed data collection forms and taking whatever corrective action is needed to ensure that the data are of high quality.

Editing is a sort of quality control check that is performed when the data are still in raw form. It is essential for detecting and correcting data collection problems before it is too late. To illustrate, the problem that occurred in the life-style research study could have been minimized and perhaps completely solved by a preliminary edit, also known as a field edit.

Definition. A ***preliminary,*** or ***field, edit*** is a quick examination of completed data collection forms, usually on the same day they are filled out.

In the life-style research study even a cursory examination of completed questionnaires immediately after the researcher received them from the interviewers would have pinpointed the interviewer who was checking both male and female on some questionnaires. The errant interviewer could then have been reinstructed about the quota-sampling requirements and asked to repeat the defective interviews, using correct procedures. Even if repeating the interviews was not feasible, the mistake could have been prevented from occurring in subsequent interviews. Therefore the number of wasted interviews would have been much lower than 25.

A field edit has two objectives: to ensure that proper procedures are being followed in selecting respondents, interviewing them, and recording their responses; and to remedy fieldwork deficiencies before they turn into major problems. Obviously, *speed* is crucial for an effective field edit: It must be done while the study is still in progress—preferably at the end of each interviewing day, and especially at the end of the first day of interviewing. Indeed, in central-location, CRT-controlled telephone interviewing, some field editing can and should be done by a supervisor as the interviews are taking place.

Illustrations of typical problems a field edit is capable of revealing include *inappropriate respondents* (as in the life-style research study discussed earlier), *incomplete interviews* (e.g., one or several questions left unanswered for no apparent reason), and *illegible* or *unclear responses* (especially in the case of open-ended questions). Upon uncovering problems of this nature, the field editor, who is usually the person in charge of supervising the fieldwork, must seek an explanation from the interviewer *immediately*, when the interview is still fresh in the interviewer's mind.

A second stage of editing, known as the final edit, or office edit, is conducted after all the field-edited data collection forms are received in a central location. In CRT-controlled telephone interviewing there are typically no physical questionnaires, and the collected data are directly stored in computer memory. Final editing of such data can be done with the aid of a computer (this topic is discussed later).

Definition. A ***final,*** or ***office, edit*** involves verifying response consistency and accuracy, making necessary corrections, and deciding whether some or all parts of a data collection form should be discarded.

An office edit is more thorough than a field edit, and the task of an office editor is somewhat more complex than that of a field editor. The following cases are illustrative of the types of problems an office editor may have to deal with.

- *Case 1.* A respondent who said he was 18 years old but indicated that he had a Ph.D. when asked for his highest level of education.
- *Case 2.* On a questionnaire containing a mixture of positive and negative Likert scale items, a respondent "strongly agreed" with *all* of them.
- *Case 3.* In response to the question "What is the most expensive purchase you have made within the past one month?" three respondents gave the following answers: respondent 1, "a new car"; respondent 2, "a vacation in Hawaii"; respondent 3, "water, gas, and electricity for my household use."

In case 1 the responses to the age and education questions appear to be inconsistent. For an 18-year-old to hold a Ph.D. degree is highly unlikely (although not impossible). Therefore the office editor has to examine a number of alternative explanations for this unusual pair of responses. Was the respondent lying about his age, his education, or both? Did the interviewer make a recording error? Or was the respondent truly brilliant, in which case the answers would be accurate? Perhaps the only way to find out which of these explanations is correct is to look for clues in the respondent's answers to other questions. For instance, if the respondent had indicated that his job title was "director of R&D" and that he had 10 years of work experience, his age as recorded would seem to be in error. Under these circumstances the editor should designate the age of this respondent as a "missing value" (more on missing values later).

Case 2 involves a set of responses that are too consistent. Since the questionnaire contained a mixture of positive and negative items, a respondent who agreed with all such items was obviously being frivolous and providing invalid answers. In this case the office editor may have no alternative other than throwing the entire questionnaire out.[1]

The situation in case 3 depicts a different type of editing problem—namely, consistency or comparability of responses *across* questionnaires to the same question. While the answers given by all three respondents are legitimate, they appear to be based on different frames of reference. The major issue facing the office editor is deciding how these diverse responses should be coded (coding will be discussed in detail in the next section). For instance, should each such response be treated as a separate category, or should only tangible-product purchases (e.g., respondent 1's answer) be treated as separate categories, with all other purchases lumped under a miscellaneous category? The specific information objectives of the study play a critical role in coding responses. In fact, before commencing the editing process, the researcher should establish a detailed set of guidelines, preferably in writing, for interpreting and categorizing open-ended responses.

The specific editing problems covered in the preceding discussion are certainly not exhaustive. Exhibit 16.1 contains a more comprehensive list of problems that an editor should look for and deal with when they do occur. Notice that several of the problems listed in Exhibit 16.1 can be handled at the field-editing stage.

A few final points about editing should be discussed. First, a number of potential editing problems can be avoided through careful planning *before* fieldwork begins. Preventing ambiguous, inappropriate, or incorrect responses is better than attempting to cure data errors after they

---

[1] Reversing the wording of some items is but one way of attempting to uncover less-than-truthful respondents. Other approaches include deliberately planting certain questions (sometimes referred to as "sleeper" questions) intended to reveal whether a respondent is too consistent or not consistent enough. For a discussion of one such approach, see Robert S. DuBoff, "Produce More Accurate Projections with 'Consistency Filtering' Technique," *Marketing News,* January 6, 1984, Section 1, pp. 16–17.

Exhibit 16.1 *A List of Problems That Editing Can Help Uncover*

1. Improper field procedures.
    a. Wrong questionnaire form used.
    b. Interview inadvertently not taken.
2. Incomplete interviews.
    a. Questions not asked.
    b. Directions not followed (proper segments of the questionnaire were not administered).
3. Improperly conducted interviews.
    a. The wrong respondent interviewed (e.g., son instead of father).
    b. Questions misinterpreted by interviewer or respondent.
    c. Evidence of bias or influencing of answers.
    d. Failure to probe for adequate answers or the use of poor probes.
    e. Interviewer apparently does not understand what type of responses constitute an answer to the actual question asked; or does not understand what the objective of the question is, and thus accepts an improper frame of reference for the respondent's answer.
    f. Interviewer's illegible writing and/or style.
    g. Interviewer recorded information which identified a respondent whose anonymity should have been protected.
    h. Other evidence of need for training or instructions to be given to interviewer (e.g., failure to write down probes, wrong abbreviations, failure to follow directions)
4. Technical problems with the questionnaire or interview.
    a. Space was not provided for needed information.
    b. The presence of unanticipated or unusually frequent extreme responses to questions, indicating a possible need for rewording of certain questions.
    c. Inappropriate or unworkable interviewer instructions not detected in the pretest.
    d. The order in which questions were asked introduces confusion, resentment, or bias into the respondent's answers.
5. Respondent rapport problems.
    a. Frequent refusal to answer certain questions.
    b. Reports of abnormal termination of the interview (or presence of hostility) due to sensitive questions.
    c. Evidence that respondent and interviewer are playing the "game" of "What answer do you want me to give?"
    d. Evidence that the presence of other people in the interview situation is causing problems.
6. Consistency problems that can be isolated and reconciled.
    a. Contradictory answers (e.g., reports no savings in one section of the interview but reports interest from bank accounts in another section).
    b. Misclassification (e.g., mortgage debt improperly reported as installment debt).
    c. Impossible answers (e.g., reports paying $600 for a new Edsel in 1970—the car should have been recorded as a "used" car; or weekly income reported on the income-per-month line).
    d. Unreasonable (and probably erroneous) responses (e.g., reports borrowing $2000 for two years to buy a car but reported monthly payments multiplied by 24 months are less than $2000; or house value is reported as being $90,000 while income is $2000 per year and the respondent claims less than a high school education).

Source: John A. Sonquist, William C. Dunkelberg, *SURVEY AND OPINION RESEARCH: Procedures for Processing and Analysis,* © 1977, pp. 43–44. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, N.J.

Research in Use 16.1 *A. C. Nielsen Company*

Computerized Editing of SCANTRACK Data

NIELSEN NEW STORE QUALITY EDIT

Certain idiosyncrasies in vendor scanning systems create data problems. Nielsen statisticians have developed a series of computerized data edit routines that are run for every store each week to determine if such idiosyncrasies have affected the data. The logic of these edit routines is based on thousands of hours of comprehensive review and research of actual scanning data problems encountered over the last 6 years. For the first several weeks during which Nielsen receives a scanning tape from a retailer, the data are thoroughly checked for consistency and completeness. Nielsen's computerized New Store Quality Edit procedure provides a comprehensive analysis of all UPC data contained on the tape to ensure that the scanning data accurately represent the sales for that particular store.

To check for data completeness, the UPCs on each tape are grouped into more than 1,000 product modules. The scan sales for each module are accumulated, and modules with sales of zero are checked to determine if the scan data are missing, or if the store does not carry items in this module.

The 1,000 product modules are collapsed into 80 product categories to check further on the completeness of data from new stores. Each category's share of total scan sales is calculated. These shares are then compared with market norms calculated from other stores in the same geographical area. Categories that deviate markedly from these norms are analyzed further to determine if the deviations are caused by a scanning problem.

Finally, the total scan dollar sales on the tape are compared with the total ACV of the store, a figure which is provided by the retailer's accounting system. This ensures that the size of the store, as measured with the scanning data, is consistent with the size as measured from an outside source in order to produce a complete scan sales assessment.

These tests are conducted for as long as it takes to ensure that the scanning data received from each store are complete and accurate. Throughout the evaluation process, as questions arise, Nielsen data analysts are in contact with the retailer's data processing center, or Nielsen field personnel visit the store to examine deviations in the scanning data.

NIELSEN STORE USABILITY EDIT

Once a store passes the initial quality evaluation and is being used in the Nielsen reporting sample, each weekly tape received is subjected to a thor-

(Continued)

Research in Use 16.1 *Continued*

ough evaluation for data accuracy and completeness. The computerized Store Usability Edit procedure systematically examines all scan data on the tape by comparing current data with data from previous weeks in the same store.

The procedure begins by taking the 80 product categories mentioned earlier and further collapsing these into 10 broad "super groups" such as frozen food, health and beauty aids, etc. The total dollar sales and UPC counts for each super group are calculated and compared by group. Averages are calculated using the previous 8 weeks of scan sales from each store. If sales and UPC counts for each group fall within a certain statistical range of the average, the data for the store are automatically accepted for use in the sample. If the sales for any product group fall outside the acceptable range, diagnostic work begins to determine the cause for the unusual sales.

To assist Nielsen data analysts in investigating questionable figures, data for individual product categories comprising the super group are automatically printed. Local merchandising is checked to determine if a promotion of some other market event has affected the figures. A UPC item exception report is also produced, listing UPCs with abnormally high sales or sell prices that could be contributing to the problem. Based on this research, a decision is made on whether the out-of-range data reflect actual market activity or reflects a scanning problem.

If further analysis is required, a more complete UPC-by-UPC analysis is performed by comparing one week's questionable data with data from the previous week. Detailed item frequency distributions are constructed for each week, and statistical tests of difference are performed to determine if the basic item structure for the 2 weeks is the same. An examination of UPCs common to both weeks as well as items that appear one week and not the other helps identify data other than the 7 days of sales data studied. This detailed analysis helps identify the type of problem. Appropriate action is taken to correct the data, or a decision is made not to use it in the sample.

The Store Usability Edit System provides for a continuous comprehensive check on scan data quality for all stores used in the Nielsen sample. The computerized nature of the system makes such detailed editing of millions of records of data each week possible. Careful attention to input data quality results in a more reliable and accurate output reporting system.

Source: From "The Nielsen Researcher," 1 (2) (Spring 1987), pp. 6–8, published by Nielsen Marketing Research, Nielsen Plaza, Northbrook, IL 60062. Used by permission.

occur. Editing is not a panacea for all data quality problems, and it is a serious mistake to view it as such.

Second, when the collected data are already in computer memory—as in the case of CRT-assisted interviews—editing can be done thoroughly and efficiently. Editing tasks that are difficult or impossible to accomplish manually, especially in large-scale surveys, can be done easily through computer editing. For instance, a computer can be programmed to check for such things as whether response values are within prespecified ranges, whether responses to key questions are consistent with those to related questions, or whether a respondent's pattern of answers deviates substantially from the average pattern of answers. Problem responses and respondents can thus be brought to the office editor's attention quickly.

Third, the role editing can play in improving data quality is much more restricted in mail surveys than in personal-interview or telephone surveys. A mail survey researcher has little control over data collection once the questionnaires are mailed out. Therefore, the only editing possible in mail surveys is a limited office edit.

Finally, the process of editing is not limited to evaluating data collected through questionnaires. It can also be used to check the quality of data collected through observation. For example, the A. C. Nielsen Company offers an information service called ScanTrack, which is based on retail sales data collected electronically from a panel of grocery stores equipped with computerized optical scanners. Once a week sales data from these stores are sent to Nielsen on tape. Before a store is considered suitable for inclusion in Nielsen's panel, the data tapes obtained from it are monitored for a period of time and subjected to a rigorous, computerized editing process to ensure that the data are trustworthy. Research in Use 16.1 describes this editing process.

In most survey research projects the process of editing, especially the office edit, goes hand in hand with the process of coding, to which we now turn.

CODING

Definition. ***Coding*** broadly refers to the set of all tasks associated with transforming edited responses into a form that is ready for analysis.

Our emphasis here will be on questionnaires used in conclusive-research projects, which invariably involve large sample sizes and computer data analyses. Exploratory-research projects are characterized by fairly informal data collection and analysis procedures. Hence a formal coding process is typically not necessary in such projects.

We can view coding as the following sequence of steps:

1. Transforming responses to each question into a set of meaningful categories.

2. Assigning numerical codes to the categories.

3. Creating a data set suitable for computer analysis.

In discussing these steps, we will use as an illustration a research study (hereafter referred to as the "survey of United States firms") conducted by the author. The primary purpose of this study was to ascertain the emphasis placed by United States firms on marketing research. A six-page questionnaire was mailed to the marketing departments of a random sample of 1000 firms in the United States. Completed questionnaires were returned by 261 firms. The first two pages of the six-page questionnaire are reproduced in Exhibit 16.2 and will be used to illustrate the various coding steps.

Transforming Responses into Meaningful Categories

How difficult and time-consuming this step is depends on the degree to which the questionnaire is structured. A structured question is precategorized—that is, it has a set of fixed-response categories. Responses to a nonstructured or open-ended question have to be grouped into a meaningful and manageable set of categories, a task that can be laborious if the respondents' answers vary widely.[2]

Consider for a moment the partial questionnaire presented in Exhibit 16.2. The questions listed are completely structured, with two exceptions: Question 4 in part I has an open-ended "other" category; question 3 in part II is open-ended. However, neither of these questions posed any major coding difficulty.

The "other" category in question 4 in part I was checked by only one firm. Had numerous firms checked this category, and had the self-descriptions provided by such firms fallen into distinct groups, splitting the "other" category into several separate categories would have been necessary. A large number of respondents checking the "other" category, however, would have strongly suggested that the question was bad—that is, the specific response categories provided were not exhaustive enough (as they should be if a question is good) to cover the bulk of the respondents. In other words, a well-designed question with an "other" category is unlikely to pose any major coding problem.

Question 3 in part II, though open-ended, sought a numerical response. Numerical responses are already coded in quantified form and need not be collapsed into a set of categories. In fact, collapsing numerical responses into a parsimonious set of categories will generally lead to a loss of information and hence is not usually advisable. In this particular case, converting the responses given in percentages (which are ratio data) into a set of rank-ordered response categories—for ex-

---

[2] An especially good discussion of the difficulties in coding answers to open-ended questions is given in Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook, *Research Methods in Social Relations* (New York: Henry Holt, 1959), Chap. 11.

Exhibit 16.2 *Two Pages of the Questionnaire Used in the Survey of United States Firms*

Part I: Please answer the following general questions about your firm and industry.

1. What is the total number of full-time employees in your firm?

__________ Fewer than 100
__________ 100–499
__________ 500–999
__________ 1000–1499
__________ 1500 or more

2. What is your firm's annual sales revenue?

__________ Less than $5 million
__________ $5 million to less than $10 million
__________ $10 million to less than $25 million
__________ $25 million to less than $100 million
__________ $100 million or more

3. How many different product and/or service categories does your firm currently market?

__________ Fewer than 10
__________ 10–24
__________ 25–49
__________ 50–99
__________ 100 or more

(Continued)

ample, 0% to less than 10%, 10% to less than 20%, and so on—would have lowered the measurement level of the data to ordinal.

A special problem in coding responses to open-ended as well as structured questions relates to the treatment of "don't know" responses. A "don't know" might be a legitimate response; that is, the respondent could not honestly answer the question. Or it might be an interviewing failure; that is, the respondent had an answer but for some reason did not divulge it.[3] An editor/coder must ascertain which of these two interpretations of "don't know" is correct. Unfortunately, this task is not easy,

---

[3] Hans Zeisel, *Say It with Figures* (New York: Harper & Row, 1968), Chap. 4, contains a good discussion of these two types of "don't know"s and ways of handling them.

Exhibit 16.2 *Continued*

---

4. Of the following categories, check *one* category that *best* describes the type of industry to which your firm belongs.

________ Consumer goods manufacturers

________ Industrial goods manufacturers

________ Consumer and industrial goods manufacturers

________ Retailers

________ Wholesalers/industrial distributors

________ Financial institutions

________ Firms offering services (other than financial)

________ Other ______________________
(please specify)

5. How would you rate the overall degree of competition among firms within your industry? Please circle the number that best describes your opinion:

Very intense	Intense	Moderate	Weak	Very weak
5	4	3	2	1

6. Would you say that the relative market position or competitive strength of your firm (when compared with other firms in your industry) is above average, average, or below average? Please check one of the following:

________ Above average

________ Average

________ Below Average

Part II: Questions in this part relate to marketing research activities of your firm.

1. Does your firm have at least one full-time employee whose *primary* duties/responsibilities relate to marketing research?

________ Yes ________ No

2. Does your firm have a separate marketing research department?

________ Yes ________ No

3. Approximately what percentage of your total marketing expenditures were used for marketing research expenditures (including the purchase of commerical marketing research services, if any) during the last year?

________ %

If you indicated 0%, skip to Question 5; otherwise, please continue with Question 4.

---

except in certain cases. For instance, a "don't know" answer to the question "Do you have any credit cards?" is most likely an interviewing failure. But a "don't know" answer to the question "Do you favor or oppose spending public funds to support certified abortion clinics?" may or may not be an interviewing failure.

There are no rules for the *best* way to treat "don't know" responses. One approach is to infer an actual response for each "don't know"—that is, to make an educated guess about what the answer might have been had the respondent not said "don't know"—on the basis of the answers to other questions. For example, a respondent's likely income bracket might be subjectively estimated from his or her age, education level, and occupation. However, this approach is fraught with questionable assumptions and hence is of dubious validity. A safer and more defensible approach is to simply classify the "don't know"s as a separate response category for each question. And if legitimate "don't know"s can be distinguished from those that are interviewing failures, one should report the latter separately as missing values, a topic discussed next.

Missing-Value Category
A response category used to code questions for which answers should have been obtained but were not for some reason.

A ***missing-value category***, as the name implies, is used to code questions for which answers should have been obtained but were not for some reason. A missing value can stem from a respondent's refusal to answer a question, an interviewer's failure to ask a question or record an answer, or a "don't know" that does not seem legitimate. Sound questionnaire design, tight control over fieldwork, and a thorough field edit can help reduce, but not necessarily eliminate, the occurrence of missing values. Questions plagued by a large number of missing values, however, are invariably indicative of a poorly designed questionnaire and/or shoddy fieldwork. In such a case caution must be exercised during subsequent analysis and interpretation of the data gathered.[4]

Assigning Numerical Codes

The next stage in the coding process is to assign appropriate numerical codes to responses that are not already in quantified form. The purpose of numerical coding is to facilitate computer manipulation and analysis of the responses. In Exhibit 16.2 the response categories in questions 1, 2, 3, 4, and 6 in part I and questions 1 and 2 in part II need to be assigned numerical codes.

The issue of assigning numbers to responses was covered at length in Chapter 11 and will not be discussed here. Nevertheless, we reemphasize that the scale or measurement level reflected by the numerical codes depends on the way a question is asked and the nature of the

---

[4] Several ways of handling missing values (similar to those mentioned in this chapter for handling "don't know"s) are available. For a good discussion of these approaches, see John A. Sonquist and William C. Dunkelberg, *Survey and Opinion Research: Procedures for Processing and Analysis* (Englewood Cliffs, N.J.: Prentice-Hall, 1977). This book is an excellent general source for editing and coding procedures.

Table 16.1 *Measurement Levels of the Responses to the Questions Used in the Survey of United States Firms*

Question Number	Construct Measured by Question	Measurement Level of Response
Part I, question 1	Number of employees	Ordinal
Part I, question 2	Annual sales revenue	Ordinal
Part I, question 3	Number of product/service categories	Ordinal
Part I, question 4	Industry type	Nominal
Part I, question 5	Industry competition intensity	Ordinal; perhaps can be assumed to be interval
Part I, question 6	Firm's competitive position	Ordinal; perhaps can be assumed to be interval
Part II, question 1	Full-time employee in charge of marketing research	Nominal
Part II, question 2	Separate marketing research department	Nominal
Part II, question 3	Percent allocation to marketing research	Ratio

variable it is attempting to measure. Table 16.1 summarizes the measurement levels achieved by responses to the various questions in Exhibit 16.2. The researcher must keep these measurement levels in mind while analyzing and interpreting the quantified responses.

Exhibit 16.3 shows one of several coding sheets used to record the data generated in the survey of United States firms (coding sheets facilitate the process of data entry into a computer but are not used when the data are entered directly into computer memory, as in the case of CRT-controlled interviewing). Each row of numbers on this coding sheet represents all the answers given by one respondent. In this study each row consisted of 80 columns (or spaces). Only 50 columns of an 80-column row were needed to summarize the answers of each respondent (notice that columns 51 through 80 are blank in Exhibit 16.3). However, when longer questionnaires are used, several rows on the coding sheet may be needed to code the answers of each respondent.

A detailed set of instructions, preferably in writing, is essential if one is to avoid errors in coding data. Table 16.2 lists the coding instructions pertaining to the partial questionnaire shown in Exhibit 16.2 and the coding sheet in Exhibit 16.3.

To see the interrelationships among the survey questionnaire (Exhibit 16.2), the coding instructions (Table 16.2), and the coding sheet (Exhibit 16.3), consider firm 50, and try to infer its responses from the numerical

Exhibit 16.3 *Illustrative Coding Sheet Used in the Survey of United States Firms*

```
        10        20        30        40        50        60        70        80
026 11243311            1010            4 7 124858
027 34455211            1010335757777736465 222626
028 12455211000         11114657577747263662112424
029 12554311000         11112222236455376466120331
030 11344211000         11113566567554256
031 5531522200705002502510223455466454253346242030
032 11254311000         11116566777766266546420909
033 24154211000         11115277766777264524120414
034 3412332100109000500511115   566632277556442940
035 11534211000         11115777777776254667511262
036 12114311000         10206356666753336444120511
037 2412432102502505502011212347577774256666441729
038 12184311000         10104567777677365557110101
039 11 45311            1111           7252  14045
040 2525421100508001001011221147477761375664121833
041 24145311000         11113355555553273657510830
042 2222422101507501501011223356366764364767311535
043 22112211000         11113236636774176337113232
044 2455432100503506000511214175567756334416122730
045 24415321000         11117277777777274767320616
046 12344311000         1010 566665677267666232333
047 111733110020200400401111          264667112531
048 11155211000         11114147477774274662112035
049 13555211000         11114454577646276767381925
050 11244321000         22103355566744374667390204
```

Note: The numbers in the first row of the coding sheet merely indicate column numbers.

Table 16.2 *Coding Instructions for the Questionnaire Used in the Survey of United States Firms*

Column Numbers on Coding Sheet	Description of Item to be Coded	Question Number	Variable Name	Range of Permissible Numerical Codes*
01–03	ID number for responding firm	—	—	001–261
04		—	—	—
05	Number of employees	Part I, question 1	V1	1–5 (5 = 1500 or more)
06	Annual sales revenue	Part I, question 2	V2	1–5 (5 = $100MM or more)
07	Number of product/service categories	Part I, question 3	V3	1–5 (5 = 100 or more)
08	Industry type	Part I, question 4	V4	1–8 (8 = other)
09	Industry competition intensity	Part I, question 5	V5	1–5 (5 = very intense)
10	Firm's competitive position	Part I, question 6	V6	1–3 (3 = above average)
11	Existence of full-time employee in charge of marketing research	Part II, question 1	V7	1–2 (2 = yes)
12	Existence of separate marketing research department	Part II, question 2	V8	1–2 (2 = yes)
13–15	Percent allocation to marketing research	Part II, question 3	V9	000–100 (*Note*: Round fractional percentages exceeding 10% to nearest whole number—e.g., 16.8% should be coded as 017. Round fractional percentages below 10% to one or two decimal places—e.g., 1.25% should be coded as 1.3; .175% should be coded as .18)

*When any variable has a missing value, leave the corresponding coding sheet columns blank.

codes in the coding sheet. You should see that firm 50 had fewer than 100 full-time employees (column 5); had annual sales revenue of less than $5 million (column 6); marketed between 10 and 24 product/service categories (column 7); was a retailer (column 8); operated in an industry characterized by intense competition (column 9); was competitively above average (column 10); had at least one full-time employee with marketing research responsibilities (column 11); had no separate marketing research department (column 12); and allocated little or none of its previous year's marketing expenditures specifically for marketing research (column 13–15).

Two features pertaining to Table 16.2 are noteworthy. First, notice that each survey question has just one corresponding variable; that is, responses to each question translate into data on only a single variable. The reason is that each question in the survey of U.S. firms had one, and *only one* possible response. Second, the entries under the Variable Name column were the symbols used in the survey of U.S. firms to identify the respective variables during computer analysis. Since variable names are no more than identification labels, they can be chosen to fit a researcher's preferences, as long as they do not violate any restrictions imposed by the specific computer program to be used.

The one-question/one-variable feature mentioned previously is not universal. A question allowing *multiple* responses will need as many variables as there are possible responses. The following question is a case in point:

Which of the following countries have you visited during the past 12 months?

_________ Canada

_________ England

_________ France

_________ Germany

_________ Japan

_________ Mexico

As few as zero or as many as six categories can be checked in response to this question. Simply assigning a series of numerical codes to the six response categories (say, 1 through 6) and treating them as pertaining to a single variable will be inadequate for coding purposes. It will be inadequate because this question is, in effect, asking respondents *six* separate questions: "Did you visit Canada during the past 12 months?" "Did you visit England during the past 12 months?" And so on. Therefore six variables, each relating to a specific country and having two possible values—for example, 1 = "no" and 2 = "yes"—will be needed. Furthermore, six columns must be set aside in the coding sheets (and/

or computer data records) to record responses to this question.[5] Similarly, "rank-order" questions such as the one that follows require as many variables (and columns) as there are objects to be ranked.

> Please rank the following fast-food restaurants by placing a 1 beside the restaurant you think is best overall, a 2 beside the restaurant you think is second best, and so on.
>
> ________ Burger King
>
> ________ McDonald's
>
> ________ Wendy's
>
> ________ Whataburger

Creating a Data Set

Definition. A ***data set***, also known as a ***data file***, is an organized collection of data records.

Definition. A ***data record*** is a series of spaces in a data storage medium capable of accommodating a prespecified number of numerical digits.

Up until the late seventies these records were primarily in the form of punched cards. At the present time, however, data are invariably stored on computer disks or tapes, which are more efficient data storage units and are eaiser to handle than a bulky deck of punched cards. Irrespective of whether the data input into the computer is through cards, optical-scan sheets, disks, or tapes, the data are arranged as a series of fixed-length records.

Case (Observation)
A sample unit for which data are available within a data set.

Each sample unit for which data are available within a data set is typically called a ***case***, or ***observation***. When the sample size is n (i.e., the number of observations $= n$) and the total number of variables embedded in the questionnaire is m, the resulting data set can be conceptually viewed as an $n \times m$ matrix of numbers, as illustrated by Table 16.3.

When coded data records are transferred to a computer, they are stored in computer memory in a format similar to the format in Table 16.3. Each x_{ij} in the table is a number and represents the coded response for the ith sample unit on the jth variable. Data analysis can begin once a data set is available in computer memory.

A researcher may sometimes want to modify the data on certain variables to create new variables for inclusion in the data analysis. To understand this process, let us once again look at Exhibit 16.2. Each question in Exhibit 16.2 represents a distinct variable. Question 1 in part

---

[5] A number of statistical software packages offer special coding instructions and analysis options for handling questions that permit multiple responses. For instance, the well-known SPSS-X package has a "MULT RESPONSE" routine for analyzing data generated by multiple-response questions.

Table 16.3 *Structure of a Data Set*

	Variables						
Observations	**1**	**2**	⋯	***j***	⋯	***m***
1	x_{11}	x_{12}	⋯	x_{1j}	⋯	x_{1m}
2	x_{21}	x_{22}	⋯	x_{2j}	⋯	x_{2m}
⋮	⋮	⋮		⋮		⋮
i	x_{i1}	x_{i2}	⋯	x_{ij}	⋯	x_{im}
⋮	⋮	⋮		⋮		⋮
n	x_{n1}	x_{n2}		x_{nj}		x_{nm}

I, for instance, represents a firm's size as measured by the number of full-time employees. Question 2 in part I also represents a firm's size, but as measured by annual sales revenue. Variables like these are directly defined by the questionnaire data and hence can be labeled ***raw variables***. New variables constructed from data on raw variables are called ***transformed variables*** (sometimes also called ***recoded variables***).

Raw Variable *A variable that is directly measured by the data in a questionnaire.*

Transformed (Recoded) Variable *A variable that is created by combining or modifying data on one or more raw variables.*

To illustrate, suppose we want to compare the results of the survey of United States firms across the following two subsamples: firms with at least 500 full-time employees *and* an annual sales revenue of at least \$25 million, and all other firms. To sort the survey results according to these subsamples, we must create a new dichotomous variable, say, X, with values 1 or 2, such that $X = 1$ for each firm whose coded response to question 1 is 3, 4, or 5 *and* whose coded response to question 2 is 4 or 5; and $X = 2$ for all other firms. The variable X is a transformation of the raw variables implied by questions 1 and 2.

To further illustrate transformed variables, suppose the following rating scales appeared in a survey of ABC Store's customers:

> Please indicate your impression of ABC Store on each of the following attributes on a scale of 1 ("Terrible") to 6 ("Excellent").

	Terrible					Excellent
1. Operating hours	1	2	3	4	5	6
2. Location	1	2	3	4	5	6
3. Layout	1	2	3	4	5	6
4. Employees	1	2	3	4	5	6
5. Quality of products	1	2	3	4	5	6
6. Variety of products	1	2	3	4	5	6

Coding the ratings on these attributes would require six raw variables—say, R1, R2, R3, R4, R5, and R6. One approach for ascertaining customers' overall impressions about ABC Store consists of the following steps:

1. Compute a total score—say, TS—for each individual customer by summing his or her ratings on the six attributes. That is,

$$TS = R1 + R2 + R3 + R4 + R5 + R6$$

2. Compute the mean value of TS across all customers to obtain their overall impressions.

In this approach TS is a transformed or recoded variable. Transformed variables such as TS, and X in the preceding illustration, are a legitimate component of data analysis. They are constructed and used in subsequent data analyses to gain more useful insights from the collected data.

Variable transformations are typically accomplished by a computer after data on raw variables have been coded and stored in computer memory. Virtually all standard computer programs are capable of creating transformed variables, as needed, just prior to data analysis. However, the researcher must specify the necessary transformed variables and instruct the computer to create them. The creation of transformed variables will increase the size of the data set. For example, if two transformed variables are created by using the data set in Table 16.3, two more columns of numbers will be added to it. Thus an expanded data set in the form of an $n \times (m + 2)$ matrix will result.

One final comment before we move on to the next section: While the chapter material so far has described editing and coding in some detail, it does not constitute a comprehensive treatment of those topics. The details of tasks like designing coding forms, hiring qualified editors/coders, and verifying their work can be found elsewhere.[6]

PRELIMINARY DATA ANALYSIS

Before analyzing a data set by using sophisticated techniques, a researcher should get a feeling for what the data are like. The purpose of preliminary data analysis is to reveal features of the basic composition of the data that have been collected. Preliminary analysis may also provide useful insights pertaining to the research objectives and suggest meaningful approaches for further analysis of the data.

Preliminary data analysis involves examining the *central tendency* and the *dispersion* of the data on each variable in the data set. The measure-

---

[6] See for example, Charles S. Mayer, "Quality Control," in *Handbook of Marketing Research*, ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 2–160 to 2–177; in the same handbook, see also Philip S. Sidel, "Coding," pp. 2–178 to 2–199.

Table 16.4 *Measures of Central Tendency and Dispersion for Different Types of Variables*

Measurement Level of Data Pertaining to Variable	Measures of Central Tendency	Measures of Dispersion
Nominal	Mode: Most frequently occurring response	Strictly speaking, the concept of dispersion is not meaningful for nominal data. But an idea about the distribution of responses can be obtained by examining their relative frequencies of occurrence.
Ordinal	Median: 50th percentile response	Range: Defined by the highest and lowest response values Interquartile range: Difference between the 75th and 25th percentile responses
Interval	Mean: Arithmetic average of response values	Standard deviation: As defined in Chapter 11
Ratio	Mean: Arithmetic average of response values	Standard deviation: As defined in Chapter 11

ment level of a variable—that is, whether the variable is nominal, ordinal, interval, or ratio—has a bearing on which measures of central tendency and dispersion will be appropriate for it. Table 16.4 summarizes the commonly used measures for different types of variables.

Note that Table 16.4 lists the same measures of central tendency and dispersion for interval and ratio data. Almost no statistical techniques have been developed specifically for ratio data; therefore they are invariably analyzed by using techniques developed for interval data. Ratio data, of course, have all the properties of interval data. Furthermore, the measures listed in Table 16.4 for nominal and ordinal data can be computed for interval and ratio data as well. Similarly, nominal data measures are also permissible for ordinal data.

A simple way of uncovering the central tendency and/or dispersion of data for virtually any variable is to construct a one-way table for it. This topic is discussed next.

Table 16.5 *Some One-Way Tables for the Survey of United States Firms*

Variable Name:
Number of Employees (V1)

Coded Response Category	Frequency of Responses	Percent of Responses	Cumulative Percent of Responses
1	128	49	49
2	78	30	79
3	17	6	85
4	7	3	88
5	31	12	100
	261	100%	

Variable Name:
Industry Competition Intensity (V5)

Coded Response Category	Frequency of Responses	Percent of Responses	Cumulative Percent of Responses
1	1	—	—
2	3	1	2
3	45	18	19
4	120	46	65
5	92	35	100
	261	100%	

Variable Name:
Full-Time Employee in Charge of Marketing Research (V7)

Coded Response Category	Frequency of Responses	Percent of Responses	Cumulative Percent of Responses
1	202	77	77
2	59	23	100
	261	100%	

One-Way Tabulation

Definition. A ***one-way table*** is a table showing the distribution of data pertaining to categories of a single variable.

Virtually all computer analysis packages are capable of generating the frequency distribution for any variable in a data set. Therefore, once a data set is in computer memory, constructing one-way tables is merely a matter of instructing the computer to do so, using a few simple commands. Table 16.5 contains the one-way tables for four of the variables in the survey of United States firms: number of employees (V1, ordinal-scaled); industry competition intensity (V5, interval-scaled); existence of

Table 16.5 *Some One-Way Tables for the Survey of the United States Firms (continued)*

Variable Name: Percent Allocation to Marketing Research (V9)			
Coded Response Category	**Frequency of Responses**	**Percent of Responses**	**Cumulative Percent of Responses**
0.00	143	60	60
0.01	3	1	61
0.02	2	1	62
0.05	1	—	62
0.10	2	1	63
0.16	1	—	63
0.20	1	—	63
0.50	4	2	65
0.90	2	1	66
1.00	16	8	74
1.50	1	—	74
2.00	13	6	80
3.00	5	2	82
4.00	4	2	84
5.00	17	8	92
7.00	1	—	92
8.00	2	1	93
9.50	1	—	93
10.00	8	4	97
13.00	1	—	97
15.00	2	1	98
20.00	3	1	99
25.00	1	—	99
30.00	1	—	99
40.00	2	1	100
	237	100%	
Missing values	24		
	261		

Note: Percentages shown in the third column have been rounded off and slightly adjusted to make them add to 100%. Also, the numbers shown under "Coded Response Category" in the one-way table for V9 do not have the same format (although they do have the same values) as the responses for V9 coded in columns 13–15 of the coding sheet shown in Exhibit 16.3. For example, the codes 000, 007, 025, and 0.2 in Exhibit 16.3 are denoted in this table as 0.00, 7.00, 25.00, and 0.20, respectively (all numbers are shown to two decimal places in this table for the sake of consistency and visual clarity).

full-time employee in charge of marketing research (V7, nominal-scaled); and percent allocation to marketing research (V9, ratio-scaled).

One-way tables are particularly appropriate for examining data on nominal- and ordinal-scaled variables since they normally only have a limited set of discrete-response categories. One-way tables can also be revealing when constructed for interval- and ratio-scaled variables (e.g., V5 and V9 in Table 16.5).

One-way tables, in addition to revealing the general nature of the data, offer some very specific benefits. First, they are helpful in *detecting certain types of coding errors.* Human errors can occur at various stages of the coding process. For instance, in the survey of United States firms an initial tabulation of data on variable V1 (number of employees) revealed two responses under category 8, although the variable should have had only five categories, numbered 1 through 5. Tracing this error back through the coding sheets and questionnaires showed that response category 3 checked by two respondents had been erroneously coded as 8. Thus, in general, preliminary one-way tabulation can facilitate data cleaning (as it is sometimes called) by pointing out glaring coding errors. However, less obvious errors—for example, mistakenly coding a 2 as a 3 when 3 is also a legitimate response category, as in the case of V1—cannot be detected by examining one-way tables.

Second, one-way tables *can provide valuable insights through comparisons with other relevant distributions.* They can be especially helpful in obtaining an understanding of the composition of respondents and looking for evidence of nonresponse error. For instance, comparing the frequency distribution of respondents on key demographic variables with appropriate distributions for the population as a whole or for nonrespondents will indicate how representative the collected data are. One such comparison was described in Chapter 15 in the discussion of nonresponse error (see Table 15.2).

Another useful technique is to compare one-way tables across similar variables within a given survey or across the same variable measured at different times. To illustrate, consider a newspaper readership study that asked respondents to indicate how often they read various syndicated columns in the newspaper. The response frequency distributions for three of the newspaper columns are presented in Table 16.6.

Table 16.6 *Some One-Way Tables for the Newspaper Readership Study*

	RESPONSE DISTRIBUTIONS					
	"Dear Abby" Column		Russ Smith's Column		Pat Buchanan's Column	
RESPONSE CATEGORY	Number	Percent	Number	Percent	Number	Percent
Usually read	228	76	108	36	42	14
Occasionally read	45	15	60	20	66	22
Rarely or never read	27	9	132	44	192	64
	300	100%	300	100%	300	100%

Mere visual inspection of the one-way tabulations for the three columns is quite revealing. The "Dear Abby" column is a clear winner, followed by Russ Smith's column and Pat Buchanan's column, in that order.

Third, one-way tables *can suggest potentially useful variable transformations.* Consider the one-way tables for variables V1 and V9 in Table 16.5. In both cases a large proportion of the responses are piled up toward one end of the range of data obtained. In other words, variables V1 and V9 have ***skewed response distributions***.[7] Variables with skewed response distributions can be troublesome in that they may not satisfy all the requirements of statistical-analysis techniques to be used subsequently. Such variables may have to be transformed into new variables that more closely fit those requirements.

Skewed Response Distribution *A one-way table in which a large proportion of responses are piled up toward one end of the range of data obtained.*

To illustrate transformations, let us focus for a moment on V1, a firm's size as measured by number of employees. One of the objectives of the survey of United States firms was to see whether a firm's size was associated with the degree of emphasis the firm placed on marketing research. The appropriate analysis technique for accomplishing this objective is chi-square contingency analysis (this technique is discussed in Chapter 18). Chi-square analysis requires a sufficient number of responses in each response category. Variable V1 was deficient in this regard since it had relatively few responses in categories 3, 4, and 5. The nature of its one-way table suggested collapsing the last three categories into a single category. The resulting transformed variable had just three categories (fewer than 100 employees, 100 to 499 employees, and 500 or more employees), with each containing an adequate number of responses.

An examination of the one-way table for V9 also resulted in the creation of a transformed variable. A majority of the responses for this variable fell into the 0% category, with the rest sparsely distributed over a broad range. Hence the researcher decided to look for the distinguishing features of firms that allocated *some* money to marketing research (irrespective of the amount) and compare them with firms that allocated *none*. For this purpose a corresponding two-category transformed variable was created.

In summary, one-way tabulations, while not as sophisticated as other data analysis techniques, can be just as insightful. Indeed, using complex techniques to analyze data without first understanding what the data look like may produce meaningless results and lead to erroneous conclusions. Even such a simple thing as computing the mean value for a variable, without having a good idea about the variable's response distribution, can be meaningless.

---

[7] The skewness of a response distribution can be regarded as the extent to which it departs from the structure of a typical bell-shaped distribution. Several ways of quantifying the extent of skewness are available. A good source of further discussion of skewness is Herman J. Loether and Donald G. McTavish, *Descriptive Statistics for Sociologists: An Introduction* (Boston: Allyn and Bacon, 1974), Chap. 5.

Variable V9 in the survey of United States firms, representing the percentage of marketing expenditures allocated to marketing research, is a case in point. The mean value for V9 is 2.2%, implying that a typical firm allocated a little over $2 for marketing research out of every $100 spent on marketing in general. But this inference is not meaningful if we look at the response distribution for V9. The mode and the median for this variable are both 0%, implying that a typical firm spent nothing on marketing research.[8] When we consider only the 94 firms that spent some money on marketing research, the mean allocation is 5.5%. But even within this subsample of firms, the median allocation is only 2%, implying that the 5.5% mean value does not truly reflect central tendency. In short, a clear understanding of the distribution of responses can help a researcher avoid erroneous inferences.

The following example, cited by Robert L. Lavidge, president of Elrick and Lavidge, Inc., a marketing research firm, further emphasizes the danger inherent in ignoring the nature of the distribution of data.

> Consider the test of a new sauce product. On the average, consumers wanted it neither really hot nor really mild. The mean rating of the test participants was quite close to the middle of the scale which had "very mild" and "very hot" as its bipolar adjectives. This happened to fit the client's preconceived notion. However, examination of the distribution of the ratings revealed the existence of a large proportion of consumers who wanted the sauce to be mild and an equally large proportion who wanted it to be hot. Relatively few wanted the in-between product, which would have been suggested by looking at the mean rating alone.[9]

Two-Way Tabulation

The objectives of most research studies include an examination of relationships among key variables. For instance, one crucial issue in the survey of United States firms was whether a firm's size was associated with how much money it allocated to marketing research. Two-way tabulation is a useful preliminary step in understanding the nature of the association between a pair of variables.

Definition. A ***two-way table*** is a table showing the number of responses in each category of one variable falling into the categories of a second variable.

---

[8] The three measures of central tendency—mode, median, and mean—will be identical only if the response distribution is bell-shaped and symmetrical. The dangers of ignoring this fact are discussed humorously and effectively, with a number of examples, by Darrell Huff, *How to Lie with Statistics* (New York: Norton, 1954), Chap. 2.

[9] "How to Keep Well-Intentioned Research from Misleading New-Product Planners," *Marketing News*, January 6, 1984, Section 2, p. 8.

Table 16.7 *Association Between Firm Size and Allocation to Marketing Research*

	Allocation to Marketing Research		
Number of Full-Time Employees	**Nothing**	**Some Funds**	**All Firms**
Less than 100	92 (77%)	27 (23%)	119 (100%)
100 to 499	41 (57%)	31 (43%)	72 (100%)
500 and over	10 (22%)	36 (78%)	46 (100%)
All firms	143	94	237* (total sample)

Note: Numbers without parentheses are the actual number of firms in each cell. The percentage figures in parentheses listed beside each of these numbers are based on the corresponding row totals.

*This value is less than 261 because of the 24 missing values for the "marketing research allocation" variable.

For two-way tabulation to be meaningful, the data on each variable must be coded into a fixed set of categories, and the number of categories should not be large. Therefore two-way tables are particularly appropriate for categorical (i.e., nominal- or ordinal-scaled) variables. Of course, two-way tables are also appropriate for interval- or ratio-scaled variables that have been transformed into ordinal-scaled variables with a limited number of categories.

Cross-Tabulation *A process involving simultaneous tabulation of data on two or more variables.*

Constructing a two-way table involves breaking down the number of responses in each category of one variable into the categories of the second variable. This process is the simplest form of ***cross-tabulation***, which refers to simultaneous tabulation of data on two or more variables. Standard computer programs capable of cross-tabulating data on any combination of variables in a data set are available for virtually all computers.

Table 16.7 summarizes the results of cross-tabulating two transformed variables in the survey of United States firms: a three-category "size" variable (less than 100 employees, 100 to 499 employees, and 500 or more employees); and a two-category "marketing research allocation" variable (firms allocating nothing to marketing research, and those allocating some funds to marketing research). Several key features of Table 16.7 deserve to be highlighted.

First, notice that the two-way table is based on a total sample of only 237 firms. This total occurs because 24 of the 261 responding firms did not answer the question dealing with allocation to marketing research, although all of them indicated the number of employees they had. An important general caveat underlies this feature—namely, the total sample size available for cross-tabulating two variables is constrained by the variable with the larger number of missing values (this maximum, effective sample size may be further reduced by missing values on the other variable). Therefore, even if just one of the two variables has a large number of missing values, the resulting two-way table may be misleading. Under such circumstances one should exercise caution in interpreting the table, especially if respondents with missing values on the variable in question are likely to differ markedly from other respondents.

Second, as shown in Table 16.7, we customarily report the responses falling within each cell as raw frequencies as well as percentages. Percentages are easy to compare and hence are helpful in gaining insight into the relationship between the variables being cross-tabulated. A simple glance at the cell percentages in Table 16.7 reveals that 78% of the firms in the "500 and over" size category, 43% of the firms in the "100 to 499" category, and 23% of the firms in the "less than 100" category had allocated some funds to marketing research. Thus the larger a firm, the more likely it is to have allocated some funds to marketing research. In other words, there is a direct, or positive, association between a firm's size and whether it allocated any money to marketing research.

Third, the cell frequencies in Table 16.7 are expressed as percentages of row totals, rather than column totals. An important question here is, In which direction (row or column) is it most meaningful to compute percentages for the cell frequencies? A general, intuitively appealing guideline is *to compute percentages in the direction of the presumed causal variable.*[10] In other words, when we can logically view one of the two variables being cross-tabulated as a *possible cause* of the other, we should base the percentages on the response totals in corresponding categories of the causal variable. Note the emphasis on "possible cause." As pointed out earlier in this book (especially in Chapter 4), cross-tabulation of data gathered through descriptive research can only suggest, but not prove, cause and effect between variables.

Let us examine Table 16.7 in light of this guideline. To view a firm's size as influencing its allocation to marketing research, rather than the other way around, seems more logical. Therefore Table 16.7 was constructed to reflect row percentages with the probable causal variable on the left and the effect variable on the top.

The general guideline for computing percentages in two-way tables has one potential problem, namely, which of the two variables is the likely causal variable may not always be clear. For instance, in the survey of United States firms, consider the cross-tabulation of data on firms' competitive positions (variable V6 in Table 16.2) and their allocations to marketing research. Whether a firm is competitively weak, average, or strong may have a bearing on how much it spends on marketing research (e.g., weaker firms may feel they do not have as much to spend and hence may allocate little or nothing to marketing research). In contrast, a firm's expenditures on marketing research may influence its relative competitve strength (e.g., firms may be competitively weak *because* of their failure to spend enough on marketing research). Thus the general guideline will not be helpful when there is no clear basis for presuming a particular direction of causality between the variables. In such cases a researcher should compute and report percentages in both directions.

---

[10] For a comprehensive discussion of this issue, see Hans Zeisel, *Say It with Figures* (New York: Harper & Row, 1968), Chap. 3; see also Loether and McTavish, *Descriptive Statistics for Sociologists,* pp. 177–182. Huff, *How to Lie with Statistics,* highlights several pitfalls related to percentages in general.

Precautions in Interpreting Two-Way Tables

Two-way tabulation, though helpful in uncovering relationships, has a few pitfalls that can easily lead a careless researcher to unwarranted conclusions. A frequent temptation when a two-way table shows evidence of a relationship, especially when one of the two variables is presumed to influence the other, is to view it as conclusive evidence of a *causal* relationship. This temptation must be resisted, however, since to infer causation from association is risky unless such evidence stems from a controlled experimental study. Two-way tables can, at best, only *suggest* the possibility of a causal relationship.

Another caveat in interpreting two-way tables is to watch out for small cell sizes and to be wary of percentages that are unaccompanied by the raw totals on which they are based. Consider the following two-way table based on a survey of 200 hospital patients:

	Patients Who	
	Jog	**Do Not Jog**
Patients with heart disease	20%	40%
Patients without heart disease	80%	60%
	100%	100%

This table appears to confirm what is commonly believed to be true: Joggers are less likely than nonjoggers to contract heart disease. Specifically, the table shows that joggers are only half as likely (20% vs. 40%) to contract heart disease. But assume for a moment that you have no prior knowledge of a possible link between jogging and lowered risk of contracting heart disease. Would you now infer that joggers are half as likely as nonjoggers to contract heart disease? You should not, not before ascertaining how many in the sample of 200 were joggers. For instance, even if there were only 5 joggers in the sample only one of whom had heart disease, the percentages in the first column would be correct. But you could not put much trust in those figures. So before you jump to conclusions about results from two-way tables (especially when the evidence is consistent with your preconceived notions), ensure that the individual cell sizes, in addition to the overall sample size, are sufficiently large.

Finally, and perhaps most important, two-way tables, by definition, are constructed by using data on just two variables at a time. While this feature makes the tables easy to interpret, it also increases the risk of drawing erroneous inferences, because the relationship between two variables may often depend on *other* variables. For example, the nature and the extent of association between students' intelligence levels and their test grades in a course may depend on factors like course content, effort exerted by students, and test format. Making inferences about the association between intelligence levels and test grades without taking into account such outside factors is risky.

In situations where a pair of variables being cross-tabulated may be influenced by other variables, the evidence presented by the two-way table may just tell part of the story. This point is an important one, but it is frequently overlooked. We will therefore discuss it in detail, using as illustration the cross-tabulations (based on hypothetical data from a survey of 1000 households) shown in Table 16.8.

Table 16.8 *Associations Between Number of Cars Owned, Household Income, and Household Size*

(a) Number of Cars Owned Versus Household Income

Household Income	Number of Cars Owned: 2 or Less	More Than 2	All Households
$25,000 or more	300 (60%)	200 (40%)	500 (100%)
Less than $25,000	200 (40%)	300 (60%)	500 (100%)

(b) Number of Cars Owned Versus Household Income and Household Size

Household Income	Household Size	Number of Cars Owned: 2 or Less	More Than 2	All Households
$25,000 or more	4 or less	260 (87%)	40 (13%)	300 (100%)
	More than 4	40 (20%)	160 (80%)	200 (100%)
Less than $25,000	4 or less	150 (60%)	100 (40%)	250 (100%)
	More than 4	50 (20%)	200 (80%)	250 (100%)

(c) Number of Cars Owned Versus Household Size

Household Size	Number of Cars Owned: 2 or Less	More Than 2	All Households
4 or less	410 (75%)	140 (25%)	550 (100%)
More than 4	90 (20%)	360 (80%)	450 (100%)

Let us examine Table 16.8(a). The evidence in this two-way table seems to suggest a negative association between a household's income and the number of cars it owns: Low income households are more likely to own more than two cars than are high-income households. But this counterintuitive finding may be due to other outside factors, such as household size, which, indeed, can have a bearing on the number of cars owned.

Table 16.8(b) is a three-way table that provides simultaneous examination of the relationships of household income and household size to the number of cars owned. It is an expanded version of Table 16.8(a) in which each cell of Table 16.8(a) is split into two cells according to household size. In Table 16.8(b), examine the two rows pertaining to households with more than 4 individuals. The percentages in these two rows are identical and suggest that *irrespective of income level,* 4 out of 5 large households are likely to own more than 2 cars. In other words, there seems to be *no association* between household income and number of cars owned for large households. Now examine the two rows pertaining to households with 4 or less individuals. The percentages in these two rows still suggest a negative association between household income and number of cars owned. While 40% of the households in the low-income category own more than 2 cars, only 13% of the households in the high-income category do so. This somewhat puzzling finding may be due to yet another variable, such as the *types* of cars owned. For example, while small households with high incomes are less likely to own more than 2 cars, the cars they do own may be more expensive than those owned by low-income households.

In short, although the three-way table, Table 16.8(b), has not completely explained the unexpected finding in the preceding two-way table, Table 16.8(a), it certainly sheds further light on the relationships being examined. Specifically, it provides the following additional insights:

1. The relationship between household income and number of cars owned is affected by household size. While there is still a negative relationship for small households, there is no relationship for large households.
2. Within each household income group, large households are more likely to own more than 2 cars than small households.

The second insight further suggests that household size, when compared with household income, may be more closely and directly associated with number of cars owned. This suggestion is supported by the two-way tabulation of data on household size and number of cars owned in Table 16.8(c).

The cross-tabulations in Table 16.8 convey an important general message: *A two-way table may not always tell the whole story about the nature of the relationship between a pair of variables.* Sometimes an apparent relationship evidenced by a two-way table may be spurious and may be explainable by incorporating other relevant variables into the tabulation. Similarly, an apparent *absence* of association between two cross-tabulated variables does not necessarily mean they are unrelated. Critical variables excluded from the table may be masking a relationship that really does exist, as illustrated by the following discussion.

Suppose a sample of 400 grocery stores is surveyed, and data are gathered for a certain brand of detergent on three variables: unit sales, promotional expenditure, and price. On the basis of these data the stores are classified as being above or below the median on each of the three variables. In other words, each store is labeled as "above the median" or "below the median" with respect to the detergent's sales, with respect to the promotional expenditure for it, and with respect to its price. The cross-tabulations of these categorical data are shown in Table 16.9.

Let us examine Table 16.9(a), the two-way tabulation of sales and promotional expenditure. This table shows no association between promotional expenditure and sales, a rather unusual finding. Since price can also influence sales, the particular set of prices used by the sample of stores may be masking the association between promotional expenditure and sales. Table 16.9(b), which is a three-way table incorporating

Table 16.9 *Associations Between Sales, Promotional Expenditure, and Price*

(a) Sales Versus Promotional Expenditure

Promotional Expenditure	**Sales** **Above the Median**	 **Below the Median**	 **All Firms**
Above the median	100 (50%)	100 (50%)	200 (100%)
Below the median	100 (50%)	100 (50%)	200 (100%)

(b) Sales Versus Promotional Expenditure and Price

Promotional Expenditure	**Price**	**Sales** **Above the Median**	 **Below the Median**	 **All Firms**
Above the median	Above the median	40 (33%)	80 (67%)	120 (100%)
	Below the median	60 (75%)	20 (25%)	80 (100%)
Below the median	Above the median	20 (25%)	60 (75%)	80 (100%)
	Below the median	80 (67%)	40 (33%)	120 (100%)

price, confirms such a masking effect. From the numbers in the first and third rows of this table (both rows relating to stores having above-median prices), we see that stores with above-median promotional expenditures are more likely (33% vs. 25%) to enjoy above-median sales than other stores. A similar conclusion is suggested by the second and fourth rows (both relating to stores with below-median prices): Stores with higher promotional expenditures are more likely (75% vs. 67%) to enjoy higher sales. In short, when price is held constant, there is a direct association between promotional expenditure and sales. Thus incorporation of price as a third variable into the original two-way table has helped reveal an important relationship that would otherwise have gone unnoticed.

The illustrations in Tables 16.8 and 16.9 imply that the larger the number of variables included in the cross-tabulated data, the smaller is the risk of making erroneous inferences. That is, three-way tables are better than two-way tables, four-way tables (having four variables cross-tabulated simultaneously) are better than three-way tables, and so on. Unfortunately, while this implication is theoretically sound, it is not always easy to implement. Even a modest increase in the number of variables included in a cross-tabulation will sharply increase the number of cells in the table. This increase will, in turn, reduce the number of sample units per cell, since the total sample size is fixed. As pointed out earlier, drawing inferences from tables with small cell sizes can be dangerous.

Furthermore, even assuming that the total sample size is large enough to warrant going beyond two-way tabulation, identifying the *most appropriate* third variable to add to the table may not always be easy. When several variables seem to be equally worthy of inclusion as the third variable, a researcher may be tempted to construct all possible three-way tables to see which variable is most useful. This approach is generally not recommended, however, since it will usually be laborious and may amount to no more than an aimless fishing expedition.

That cross-tabulations may quickly become unwieldy and meaningless underlines the need to plan ahead and map out a systematic analysis strategy well in advance. Sound exploratory research, including a thorough review of past studies related to the topic being researched, is extremely useful in developing hypotheses about the nature of the variables of interest and the relationships among them. Such hypotheses, at least in the form of informal hunches, are essential in deciding which two-way tables to construct and which of those to expand by incorporating other relevant variables. Moreover, recognizing that an apparent relationship may be spurious—or that an apparent lack of relationship may be suspect—is not possible without the benefit of prior hypotheses. In short, a researcher with no hypotheses but hoping to gain insights through cross-tabulations is like a compassless ship hoping to go from point A to point B. If the ship ever gets to point B, it will only be after a great deal of inefficiency in the form of numerous miscues and wasted resources.

SUMMARY

Editing is a quality control check on the raw data; it involves inspection and any necessary correction of completed data collection forms. A preliminary stage of editing, known as the field edit, is conducted when the data are being collected. Its purpose is to verify that interviewers are following proper procedures and to correct any fieldwork problems before it is too late. A more comprehensive edit, known as the final, or office, edit, is conducted after all the completed data collection forms are returned to a central office. Major responsibilities of an office editor include checking for response consistencies, deciding on appropriate ways to handle "don't know" responses and missing values, and classifying open-ended responses.

The process of coding immediately follows, or often overlaps with, the office edit. It includes classifying open-ended responses into categories and assigning numerical values to response categories that are not already in quantified form. Once the data are coded, a data set, or data file, can be created in computer memory just prior to data analysis. A data set is a matrix of numbers in which each row contains the coded answers given by a single respondent (typically referred to as a case or observation).

A good practice is to conduct a preliminary analysis to get an initial idea about the nature of the data. Preliminary analysis involves obtaining and examining the central tendency and the dispersion of the data for all variables (including transformed variables) of interest. The choice of the most appropriate measure of central tendency and dispersion to use depends on whether the variables are nominal, ordinal, interval, or ratio.

A rather simple way to ascertain what the data look like is a one-way tabulation, that is, the frequency distribution of responses for each variable. In addition to revealing the general nature of the data, one-way tables are helpful in detecting certain types of coding errors; comparing the distribution of data on a variable with other relevant distributions, such as the overall population distribution and the distributions of similar variables; and suggesting meaningful variable transformations.

Two-way tabulation involves breaking down the number of responses in each category of one variable into the categories of a second variable. The purpose of a two-way tabulation is to explore the extent and the nature of possible associations between pairs of variables. The numbers in a two-way table are usually reported as raw frequencies and percentages. A general guideline for computing percentages in a two-way table is to compute them in the direction of the presumed causal variable. However, if the presumed causal variable is not clear-cut, one should compute the percentages in both directions.

While two-way tables can provide valuable insights about relationships between variables, they can easily mislead an unwary researcher. Thus the following precautions are worth bearing in mind: First, two-way tables can suggest, but cannot convincingly show, cause and effect between variables. Second, two-way tables with very small cell sizes, or those that present the evidence solely in percentages, may be suspect. Third, evidence in a two-way table of a particular type of relationship, or lack thereof, between two variables is not necessarily conclusive.

Relevant variables excluded from a two-way table may lead to a spurious relationship or may mask a true relationship. When external variables are suspected of playing a role in a relationship, a researcher should construct and examine appropriate three-way, or even larger, tables. Unfortunately, inadequate sample sizes and lack of information about which external variables to include often pose practical constraints in constructing multiway tables. Moreover, numerous cross-tabulations may overwhelm a researcher. Therefore the researcher should have at least a tentative analysis plan beforehand specifying which cross-tabulations will be constructed and how detailed each will be. Formulating a clear-cut analysis plan requires adequate exploratory research capable of leading to hypotheses concerning the variables of interest.

KEY TERMS

Editing
Preliminary (field) edit
Final (office) edit
Coding
Missing-value category
Data set (file)
Data record
Case (observation)
Raw variable
Transformed (recorded) variable
One-way table
Skewed response distribution
Two-way table
Cross-tabulation

QUESTIONS

1. Briefly discuss the similarities and differences between a *field edit* and an *office edit.*
2. "Editing plays no role in acting as a quality control check on mail survey data." Discuss this statement.
3. What are the two types of "don't know" responses? Give an example of each.
4. Why do questions permitting multiple responses require special attention during coding?
5. Refer to Exhibits 16.2 and 16.3 and Table 16.2. Derive a descriptive profile of firm 45 on variables V1 through V9.
6. Define these terms: *case, data record, data set.*
7. What are *transformed variables*?
8. Transform variable V9 in the survey of United States firms into an appropriate three-category new variable (excluding the missing-value category) that reflects varying ranges of allocations to marketing research. Explain your rationale. Also, construct a one-way table for the new variable.

9. What are three potential uses of one-way tabulation?

10. What is a *skewed distribution*? Suppose two ordinal variables, each with eight categories, have extremely skewed distributions. What type of variable transformation would you create for the two variables to make their cross-tabulation meaningful? Briefly explain your answer.

11. What is the general principle for computing percentages in two-way tables? What is the one exception to this principle?

12. What are the three major precautions that should be taken in interpreting two-way tables?

13. From the information in Table 16.9, construct a two-way tabulation of the detergent's price and its unit sales. Is there a relationship between these two variables? If so, what is its nature? If not, why not?

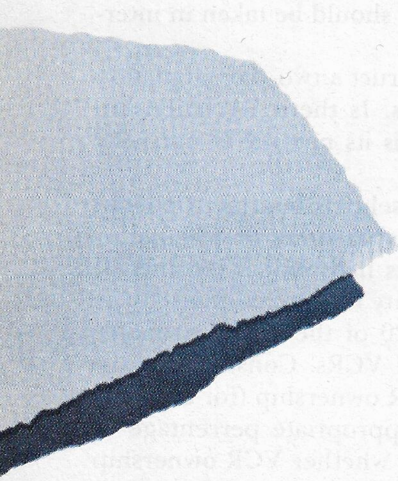
14. A survey of a random sample of 500 households in a metropolitan area revealed that 200 of them had annual incomes of $30,000 or more while the remaining 300 households had annual incomes of less than $30,000. One hundred and twenty of the 200 households with incomes of $30,000 or more and 120 of the 300 households with incomes of less than $30,000 owned VCRs. Construct a two-way table of household income versus VCR ownership (for each cell show the actual number as well as the appropriate percentage of responses). Interpret the table and discuss whether VCR ownership is associated with income levels and, if so, the nature of the association.

15. A stock brokerage company conducted a random survey of 300 of its customers. One question on the survey asked respondents whether they had been experiencing problems with any aspect of the service within the past three months. One-third of the respondents answered "yes" to this question while the remaining two-thirds answered "no." On the basis of respondents' transactions with it, the company was able to group the respondents into "high-volume" and "low-volume" categories. Of the respondents experiencing problems, 40 were "high-volume" customers; while of those not experiencing problems, 78 were "high-volume" customers. The company is interested in exploring the question of whether customers experiencing problems were likely to do less business with it than those not experiencing problems. What insight do the available data offer regarding this question? Explain your answer.

16. What are the practical constraints involved in constructing and interpreting multiway tables? Summarize the general implication of these constraints in a sentence or two.

CHAPTER 17

Hypothesis Testing & Its Role in Data Analysis



CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between descriptive analysis and inferential analysis.*
- *State the null and alternative hypotheses pertaining to a variety of decision situations requiring formal hypothesis testing.*
- *Define Type I and Type II errors and state the relationship between the two.*
- *Define significance level and power of a hypothesis test.*
- *State the general steps involved in conducting a hypothesis test.*
- *Differentiate between univariate and multivariate analysis.*
- *Indicate the main distinction between parametric and nonparametric procedures.*
- *Describe when and how to use the chi-square goodness-of-fit test and the Kolmogorov-Smirnov test.*
- *Identify and use the appropriate test pertaining to hypotheses involving a single mean, a single proportion, two means (when the two samples are independent and when they are dependent), and two proportions.*

DESCRIPTIVE VERSUS INFERENTIAL ANALYSIS

Descriptive Analysis
Preliminary data analysis (e.g., computing measures of central tendency and dispersion) that helps summarize the general nature of variables included in a study and the interrelationships among them.

Inferential Analysis
Data analysis that goes beyond descriptive analysis; it involves verifying specific statements or hypotheses about the population(s).

The data analysis procedures discussed in Chapter 16—namely, computing measures of central tendency and dispersion, as well as constructing one-way, two-way, and multiway tables—help the researcher summarize the general nature of the study variables and the interrelationships among them. Such procedures constitute ***descriptive analysis*** and can provide valuable insights, as demonstrated by the examples in the previous chapter. In fact, certain research studies may require no more than descriptive analysis of the data. In many studies, however, one must go beyond descriptive analysis in order to verify specific statements, or *hypotheses,* about the population(s) of interest. Data analysis aimed at testing specific hypotheses is usually called ***inferential analysis.*** This chapter describes the general procedure involved in hypothesis testing, discusses the role of hypothesis testing in data analysis, and outlines several hypothesis tests frequently encountered in marketing.

OVERVIEW OF HYPOTHESIS TESTING

A useful starting point for discussing hypothesis testing is to consider the following situations, which illustrate critical questions typically faced by decision makers.

Situation A. Karen, product manager for a line of apparel, is wondering whether to introduce the product line into a new market area. A recent survey of a random sample of 400 households in that market showed a mean income per household of $13,000. On the basis of past experience and of comprehensive studies in current market areas, Karen strongly feels that the product line will be adequately profitable only in markets where the mean household income (across *all* households) is greater than $12,000. Should Karen introduce the product line into the new market?

Situation B. Tom, advertising manager for a frozen-foods company, is in the process of deciding between two TV commercials, X and Y, for a new frozen food to be introduced shortly. Commercial X runs for 20 seconds, and commercial Y runs for 30 seconds. Therefore for a given number of exposures, commercial Y will be more expensive than commercial X. Tom feels that commercial Y will also be more effective in creating awareness for the new product, but he is not sure.

The two commercials were recently shown during identical TV programs in two comparable test cities (only one of the two was shown in each city). After the commercials were shown, a random sample of 200 adults was interviewed by telephone in each city. In the city in which commercial X was shown, 40 of the 200 respondents were aware of the new frozen food; that is, the awareness rate for commercial X was 20%. In the other city the awareness rate for commercial Y was 25%. Can Tom conclude that commercial Y will be more effective in the total market for the new frozen food?

What features do situations A and B have in common? Clearly, in order to reach a final decision, both Karen and Tom have to make a general inference from sample data. However, making generalizations from sample data is a feature implicit in virtually all conclusive-research projects and hence is not unique to situations A and B. As we have seen in the chapters on sampling (Chapters 12, 13, and 14), the purpose of any sampling study is to learn *something* about the population.

A more distinctive feature of situations A and B, one that is more directly relevant to hypothesis testing, is that each situation implies a criterion on which the final decision depends. In situation A the criterion is the mean income across all households in the new market area under consideration. Specifically, if the mean population household income is greater than $12,000, Karen should introduce the product line into the new market. In situation B the criterion is the relative degrees of awareness likely to be created by the two commercials in the population of all adult consumers. Specifically, Tom should conclude that commercial Y is more effective than commercial X only if the anticipated population awareness rate for Y is greater than that for X.

Stated differently, Karen's decision making in situation A is equivalent to either accepting or rejecting the following hypothesis: "The population mean household income in the new market area is greater than $12,000." Similarly, Tom's decision making in situation B is equivalent to either accepting or rejecting the following hypothesis: "The potential awareness rate that commercial Y can generate among the population of consumers is greater than that which commercial X can generate." A situation calling for formal hypothesis testing will usually stipulate a specific criterion for choosing between alternative inferences or courses of action. However, as we will see later, certain types of hypothesis tests may not have a criterion as clear-cut as the criteria in situations A and B. Furthermore, in many real-life situations final decisions may depend on several factors rather than on a single, clear-cut criterion; situations A and B have been simplified intentionally to highlight the salient features of hypothesis testing.

Null and Alternative Hypotheses

The first step, after one recognizes that a particular decision situation requires formal hypothesis testing, is to state a null hypothesis and an alternative hypothesis. We will use the symbols H_o and H_a to denote the null and alternative hypotheses, respectively. Before discussing the specific meanings of H_o and H_a, we make one general comment about stating hypotheses: Hypotheses always pertain to *population* parameters or characteristics rather than *sample* characteristics. It is the population, *not* the sample, that we want to make an inference about from limited data. Although this point may seem obvious, in the author's experience, students often get confused about it when formally stating and interpreting hypotheses.

The null and alternative hypotheses complement each other; they are mutually exclusive and collectively exhaustive. In other words, the two hypotheses are stated in such a way that H_a will not be accepted if the sample evidence strongly supports H_o. Similarly, H_a will be accepted if the sample evidence is strong enough to reject H_o. In situation A, with μ denoting the population mean household income, the two hypotheses are as follows:

$$H_o: \mu \leq \$12{,}000;$$
$$H_a: \mu > \$12{,}000.$$

You are probably wondering how one decides which hypothesis is the null (H_o) and which is the alternative (H_a). This issue depends to a certain extent on the nature of the decision to be made. However, of the two hypotheses, the null should be stated more conservatively; that is, failure to reject H_o from the sample evidence should preserve the status quo. Notice that in situation A the product line will not be introduced into the new market area (i.e., status quo will be maintained) if H_o is not rejected. If a decision situation is such that one of the two hypotheses implies a strict equality (e.g., $\mu = \$12{,}000$), then that hypothesis should always be treated as H_o.

Type I and Type II Errors

Inferences stemming from hypothesis testing are subject to error. One cannot be entirely certain whether a null hypothesis is true or false unless a study involving a census is undertaken.

In the absence of a census study, whether or not the null hypothesis should be rejected (and the alternative hypothesis accepted) is based on evidence from the sample. Such evidence can at times lead to wrong inferences about the population because of possible sampling error. Specifically, the sample evidence might suggest that H_o should be rejected when it is actually true, or that H_o should not be rejected when it is actually false. The resulting mistakes are called Type I and Type II errors, respectively.

Definition. A ***Type I error*** is committed if the null hypothesis is rejected when it is true.

Definition. A ***Type II error*** is committed if the null hypothesis is *not* rejected when it is false.

A Type I error may occur when H_o is rejected. Likewise, a Type II error may occur when H_o is not rejected. Therefore the possibility of committing a Type I error *or* a Type II error is inherent in any hypothesis-testing procedure. Although Type I and Type II errors cannot be entirely avoided, the probability of occurrence of one of them can be constrained to be within an acceptable level. Virtually all hypothesis-testing procedures place a limit on the probability of committing a Type I error. This limit is specified by the researcher or decision maker and, as we will see a little later, forms the basis for deciding when to reject H_o and when not to do so. The term *significance level* refers to the upper-bound probability of a Type I error.

Definition. The ***significance level*** associated with a hypothesis-testing procedure is the maximum probability of rejecting H_o when using that procedure and H_o is actually true.

The symbol α (Greek letter alpha) is typically used to denote the significance level. The complement of the significance level (i.e., $1 - \alpha$) is called the *confidence level.*

The symbol β (Greek letter beta) is typically used to denote the probability of committing a Type II error. The quantity $(1 - \beta)$, which is the probability of avoiding a Type II error, is called the power of the test, or simply *power.*

Definition. The ***power*** of a hypothesis test is the probability of rejecting H_o when H_o is false.

Table 17.1 provides a summary of the key concepts and terms involved in hypothesis testing. While α can be, and typically is, specified by the researcher, β is difficult to constrain because its value depends on the actual (but unknown) value of the population parameter. Moreover, there is an inverse relationship between α and β. In other words, as the probability of a Type I error is lowered, the probability of a Type II error goes up (i.e., power decreases). Therefore it is difficult to limit *both* α and β to be within prespecified levels. The relationship of β (and

Table 17.1 *Summary of Errors Involved in Hypothesis Testing*

Inference Based on Sample Data	Real State of Affairs	
	H_o Is True	**H_o Is False**
H_o Is True	Correct decision Confidence level = $1 - \alpha$	Type II error P (Type II error) = β
H_o Is False	Type I error Significance level = α*	Correct decision Power = $1 - \beta$

*Term α represents the maximum probability of committing a Type I error.

hence power) to α and the population parameter value is discussed further in the appendix to this chapter.

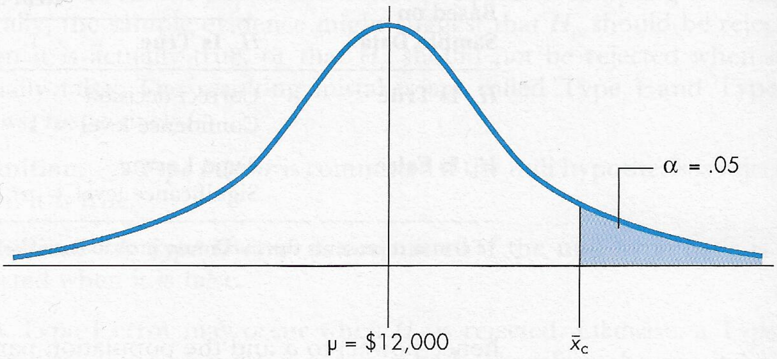
Let us now examine the role played by the significance level in deciding when to reject H_o. In situation A, suppose Karen does not want to allow more than a 5% chance of rejecting H_o when $\mu \leq \$12{,}000$ (i.e., $\alpha = .05$). In other words, she does not want to reject H_o unless the sample evidence is overwhelmingly against it. The significance level is the key in deciding what type of evidence can be considered as being "overwhelmingly against" H_o.

Recall from Chapter 12 that mean values of samples selected repeatedly from a given population will fall on either side of the population mean and form a distribution (called the sampling distribution) around it. Intuitively, sample mean ($\bar{x}$) values greater than \$12,000—that is, $\bar{x}$-values on the right-hand side of the sampling distribution centered on $\mu = \$12{,}000$—suggest that H_o may be false. And, more importantly, the farther away to the right $\bar{x}$ is, the stronger is the evidence against H_o. Therefore beyond some critical $\bar{x}$-value, say $\bar{x}_c$, Karen should be able to conclude that the evidence is overwhelmingly against H_o. Specifically, she should choose $\bar{x}_c$ such that any sample whose mean exceeds $\bar{x}_c$ has less than a 5% chance of having come from a population whose mean is \$12,000 or less. This criterion is shown graphically in Exhibit 17.1.

How does one select an appropriate significance level (or confidence level) for a hypothesis test? For instance, in situation A, how or why did Karen choose a significance level of .05? The choice of a particular significance level is subjective and depends on the nature of the problem as well as on the degree of risk the decision maker is willing to take. The following example illustrates this point.

Example. Consider two firms, ABC and XYZ, each considering introducing a new product that radically differs from its current product line. Company ABC has a well-established customer base and enjoys a distinct reputation for its existing product line. In contrast, XYZ has neither a loyal clientele nor a distinct image for its present

Exhibit 17.1 *Identifying the Critical Sample Mean Value*



products. Which of these two firms should be more cautious in making a decision to introduce the new product?

Intuitively, we feel that ABC should be more cautious, since it apparently has more to lose than XYZ by making a hasty go-ahead decision. Stated differently, ABC perhaps should set a lower significance level (or higher confidence level) than XYZ, if both firms are to make a go–no-go decision on the basis of a suitable hypothesis test using sample data pertaining to potential customer acceptance of the new product. Thus different significance levels for the two firms (e.g., $\alpha = .01$ for ABC and $\alpha = .20$ for XYZ) would be quite appropriate in this situation. Of course, the willingness of management of the two firms to assume risk also plays a role: Generally, the more risk-averse management is, the lower the significance level should be.

In summary, the choice of a suitable significance level for a hypothesis test is, or should be, situation-specific. There is no standard significance level that is equally appropriate under all circumstances. In other words, although many hypothesis tests use seemingly standard significance levels of .01, .05, or .10, there is nothing sacred about these numbers. It is not wrong to use a significance level of, say, .20 or .30, if the circumstances warrant it.

Decision Rule for Rejecting the Null Hypothesis

Definition. A ***decision rule*** is a guideline that specifies the sample evidence necessary to reject the null hypothesis.

In situation A, for instance, Karen needs a decision rule specifying a *critical* sample mean income level ($\bar{x}_c$) such that if the actual sample mean ($\bar{x}$) exceeds that level she can reject H_o. The critical value to be incorporated in the decision rule depends on the significance level (α) specified for the hypothesis test. Suppose Karen does not want to allow more than a 5% chance of rejecting H_o when the population mean is actually less than or equal to \$12,000 (i.e., Karen sets the value of α at .05). Exhibit 17.1 shows the sampling distribution curve associated with a population mean of \$12,000. Recall from Chapter 12 that this curve will have properties of the normal probability distribution since the sample size in situation A is sufficiently large ($n = 400$). Therefore the shaded area in Exhibit 17.1 corresponds to the probability of obtaining sample means in the tail portion of the curve (i.e., $\bar{x} > \bar{x}_c$) when the population mean is \$12,000. Since this probability is specified to be .05 (i.e., $\alpha = .05$), there is only a 5% chance that the population mean is \$12,000 if the sample mean obtained exceeds $\bar{x}_c$ (the chances that the population mean is less than \$12,000 would be even lower). Therefore the decision rule for Karen can be stated as: "Reject H_o if the sample mean exceeds $\bar{x}_c$."

Of course, the exact value of $\bar{x}_c$ needs to be specified before Karen can make a decision. The value of $\bar{x}_c$ can be determined by using our knowledge that the sampling distribution in this situation is a normal probability distribution.

Specifically, every mean $\bar{x}$ has a corresponding, equivalent standard normal deviate z, given by

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}, \tag{17.1}$$

where $\sigma_{\bar{x}}$ is the standard deviation associated with the sampling distribution curve. When $\sigma_{\bar{x}}$ is unknown, it is approximately estimated by $s_{\bar{x}}$, the standard error of the sample mean, given by $s_{\bar{x}} = s/\sqrt{n}$, where s is the sample standard deviation and n is the sample size. Therefore the expression for z can be modified as follows:

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}}. \tag{17.2}$$

Rearranging the terms in Eq. (17.2) gives

$$\bar{x} = \mu + z s_{\bar{x}}. \tag{17.3}$$

Substituting $\bar{x}_c$ for $\bar{x}$ and z_c for z in Eq. (17.3) yields

$$\bar{x}_c = \mu + z_c\, s_{\bar{x}}, \tag{17.4}$$

where z_c is the standard normal deviate corresponding to the critical sample mean, $\bar{x}_c$. That is, z_c is a critical z-value that corresponds to a significance level of α under the standard normal curve. The areas, or probabilities, under different sections of the standard normal curve and their corresponding z values are tabulated in Appendix B at the end of the book. The value of z_c corresponding to a specified α-level can be obtained from the table in Appendix B.

Let us return now to situation A. Suppose the standard deviation of household income for the sample of 400 households is \$8000. The standard error of the mean ($s_{\bar{x}}$) is given by

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{\$8000}{\sqrt{400}} = \frac{\$8000}{20} = \$400.$$

The critical mean household income $\bar{x}_c$ can now be computed through the following two steps:

1. Determine the critical z-value z_c such that the area to its right under the standard normal curve is α, which is .05 in this case. From Appendix B, $z_c = 1.645$.
2. Substitute the values of z_c, $s_{\bar{x}}$, and μ (under the assumption that H_o is "just" true) into Eq. (17.4) and solve for $\bar{x}_c$.

A note about the term *just true* is needed here. When H_o is stated as an inequality (as it is in situation A), it can be true for a *range* of population parameter values. It is "just" true, or "barely" true, when the population parameter is exactly equal to the hypothesized value. Furthermore, the probability of committing a Type I error is greatest when H_o is just true. Therefore calculating the value of $\bar{x}_c$ when H_o is just true guarantees that the probability of committing a Type I error will be no more than the specified significance level (α).

In situation A, H_o is just true when $\mu = \$12{,}000$. From Eq. (17.4),

$$\begin{aligned} \bar{x}_c &= \mu + z_c s_{\bar{x}} = \$12{,}000 + 1.645 \times \$400 \\ &= \$12{,}000 + \$658 = \$12{,}658. \end{aligned}$$

Karen's decision rule can be stated as follows: "If the sample mean household income is greater than \$12,658, reject the null hypothesis and introduce the product line into the new market area." The same decision rule can also be stated in terms of z-values, as follows: "If the z-value corresponding to the sample mean household income is greater than 1.645, reject the null hypothesis and introduce the product line into the new market area." In general, the latter format is used to specify the decision rule for rejecting H_o, because it is cast in terms of a standard test statistic.

Definition. The ***test statistic*** is a standard variable whose value is computed from sample data and compared with a critical value (obtained from an appropriate probability table) to decide whether or not to reject the null hypothesis.

Every hypothesis test has a corresponding test statistic, which depends on the sampling distribution involved. Since the sampling distribution in situation A has the properties of the normal probability distribution, the appropriate test statistic is the z-variable. The value of the test statistic is simply the z-value corresponding to $\bar{x} = \$13{,}000$. From Eq. (17.2),

$$z = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\$13{,}000 - \$12{,}000}{\$400}$$

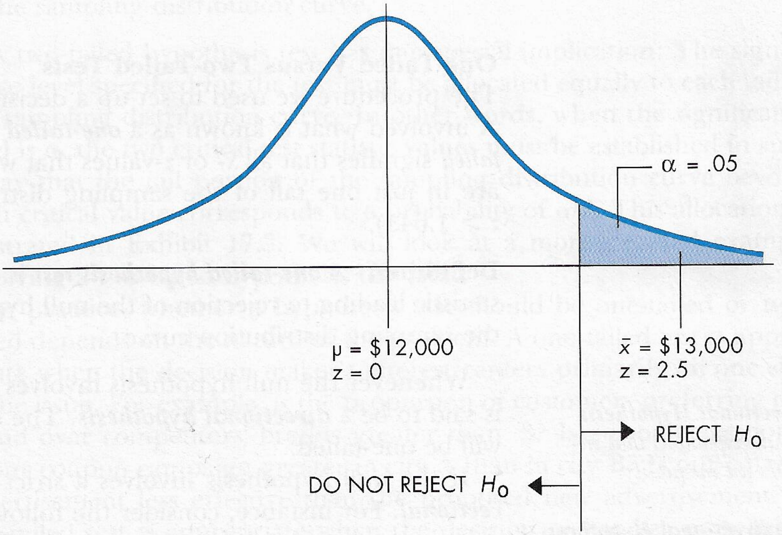
$$= \frac{\$1000}{\$400} = 2.5.$$

Clearly, Karen should reject H_o from the sample evidence since $\bar{x} > \bar{x}_c$ and $z > z_c$. In other words, she should introduce the product line into the new market area. If she does so, the probability of her committing a Type I error will be less than .05, as shown in Exhibit 17.2.

The intuitive notion underlying this hypothesis test is embedded in the following question: "What are the chances that a sample mean income as high as \$13,000 (or a test statistic value as high as 2.5) will have occurred when the true population mean income is \$12,000 or less?" The answer to this question in situation A is, "Less than 5%, the subjectively selected significance level." Therefore from the sample evidence the hypothesis that the population mean income is \$12,000 or less is not tenable and should be rejected.

The probability of obtaining an $\bar{x}$-value as high as \$13,000 or more when μ is only \$12,000 is given by the area under the distribution curve to the right of $z = 2.5$. From Appendix B this probability is only .0062.

Exhibit 17.2 *Critical Value for Rejecting the Null Hypothesis*



This value is sometimes called the *actual* significance level, or the *p-value*, of the test, to distinguish it from α, which is the preset, maximum significance level. The actual significance level of .0062 in this case means that the odds are less than 62 out of 10,000 that the sample mean income of \$13,000 would have occurred entirely due to chance (i.e., when the population mean income is \$12,000 or less).

For decision makers facing go–no-go decisions (such as whether a new product should be introduced or a TV commercial should be run) the actual significance level provides an estimate of the risk involved in erroneously making a go decision. Therefore rather than making their go–no-go decisions on the basis of a prespecified "standard" significance level (e.g., .05), they can examine the actual significance level and then decide whether they are willing to tolerate that level of risk of making an unwarranted go decision. For instance, suppose the test statistic value (i.e., z-value) was 1.0 instead of 2.5 in situation A. From Appendix B the actual significance level, or p-value, corresponding to $z = 1.0$ is .1587 (i.e., the area under the normal probability curve to the right of $z = 1.0$ is .1587). This p-value implies that there is about a 16% chance that a test statistic value as high as 1.00 would have been obtained when the population mean was only \$12,000. In other words, there is about a 16% chance that Karen would be making a mistake if she decided to introduce the product line on the basis of the sample evidence. If Karen finds this level of risk (i.e., probability of a Type I error of .16) acceptable, she may decide to go ahead and introduce the product line rather than hold back simply because the z-value of 1.0 falls short of the critical value of 1.645 at the prespecified, and arbitrary, significance level of .05.

One-Tailed Versus Two-Tailed Tests

The procedure we used to set up a decision rule for Karen in situation A involved what is known as a ***one-tailed hypothesis test***. The term *one-tailed* signifies that all $\bar{x}$- or z-values that would cause Karen to reject H_o are in just one tail of the sampling distribution (i.e., $\bar{x} > \$12,658$, or $z > 1.645$).

Definition. ***A one-tailed hypothesis test*** is one in which values of the test statistic leading to rejection of the null hypothesis fall only in one tail of the sampling distribution curve.

Directional Hypothesis
A null hypothesis that involves an inequality.

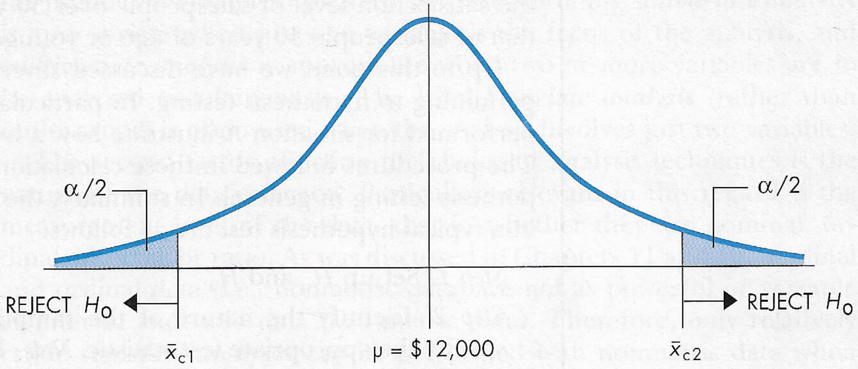
Whenever the null hypothesis involves an inequality (i.e., $\leq$ or $\geq$), it is said to be a ***directional hypothesis***. The corresponding hypothesis test will be one-tailed.

Nondirectional Hypothesis
A null hypothesis that involves a strict equality.

If the null hypothesis involves a strict equality (i.e., =), it is ***nondirectional***. For instance, consider the following pair of hypotheses:

$$H_o:\mu = \$12,000;$$
$$H_a:\mu \neq \$12,000.$$

Exhibit 17.3 *Critical Sample Mean Values for a Two-Tailed Test*



Intuitively, both very high and very low values of $\bar{x}$ should lead to rejection of H_o. Therefore the decision rule for rejecting H_o will have *two* critical $\bar{x}$-values—one below $12,000 (say $\bar{x}_{c1}$) and the other above $12,000 (say $\bar{x}_{c2}$). The decision rule will be, "Reject H_o if $\bar{x} < \bar{x}_{c1}$ *or* if $\bar{x} > \bar{x}_{c2}$." As shown in Exhibit 17.3, $\bar{x}$-values in either tail of the curve imply that H_o should be rejected.

Definition. A ***two-tailed hypothesis test*** is one in which values of the test statistic leading to rejection of the null hypothesis fall in both tails of the sampling distribution curve.

A two-tailed hypothesis test has one special implication: The significance level specified for the test must be allocated equally to each tail of the sampling distribution curve. In other words, when the significance level is α, the two critical test statistic values must be established in such a way that the tail portion of the sampling distribution curve beyond each critical value corresponds to a probability of α/2. This allocation is illustrated in Exhibit 17.3. We will look at a more detailed example involving a two-tailed hypothesis test later.

In practice, whether a hypothesis test should be one-tailed or two-tailed depends on the nature of the problem. A one-tailed test is appropriate when the decision maker's interest centers primarily on one side of the issue. For example, is the proportion of customers preferring our brand over competitors' brands greater than .3? Is customer response to our coupon campaign greater in city A than in city B? Is our current advertisement less effective than the proposed new advertisement? A two-tailed test is appropriate when the decision maker has no a priori reason to focus on one side of the issue. For example, is the average useful life for our appliance as perceived by consumers different from the objectively determined average life of 10 years? Is test market C different from test market D in terms of average household incomes? Is the satisfaction level of salespeople over 30 years of age different from that of salespeople 30 years of age or younger?

Up to this point we have discussed several key terms and concepts pertaining to hypothesis testing. In particular, the calculations we have performed for situation A illustrate how a hypothesis test is conducted. The procedures followed in those calculations are representative of hypothesis testing in general. In summary, the sequence of tasks involved in a typical hypothesis test are as follows:

Step 1. Set up H_o and H_a.

Step 2. Identify the nature of the sampling distribution curve and specify the appropriate test statistic. *Note:* In situation A the sampling distribution was the normal curve and the test statistic was the z-variable. But as we will see later, depending on the specific problem, the appropriate sampling distribution and test statistic will vary.

Step 3. Determine whether the hypothesis test is one-tailed or two-tailed.

Step 4. Taking into account the specified significance level, determine the critical value (*two* critical values for a two-tailed test) for the test statistic from the appropriate statistical table.

Step 5. State the decision rule for rejecting H_o.

Step 6. Compute the value for the test statistic from the sample data.

Step 7. Using the decision rule specified in step 5, either reject H_o or reject H_a.

We will use this sequence of steps as the basic framework in describing the various hypothesis tests to be covered subsequently.

ROLE OF HYPOTHESIS TESTING IN DATA ANALYSIS

The scenarios in situations A and B presented earlier were such that the key issue facing the decision maker in each situation could be readily stated as a formal hypothesis. However, in not all situations will the need to test specific hypotheses be that apparent. Nevertheless, virtually all data analysis techniques to be discussed in the next two chapters involve hypothesis testing, at least indirectly. In other words, hypothesis testing may be viewed as an integral (although not always formally stated) part of most analysis procedures that go beyond merely describing the nature of the data. Therefore we will review at this point certain key determinants of the appropriate analysis technique (and hence the appropriate hypothesis test) to use in a given situation.

Two factors are crucial in choosing an appropriate analysis procedure: the number of variables to be analyzed and the nature of the data collected on each variable. Analysis procedures are broadly classified as being univariate or multivariate, depending on the first factor, that is, the number of variables involved. As the terms imply, ***univariate analysis*** is appropriate when just one variable is the focus of the analysis, and ***multivariate analysis*** is appropriate when two or more variables are to be analyzed simultaneously. The label ***bivariate analysis*** (rather than multivariate) is often used when the analysis involves just two variables.

Univariate Analysis
Any procedure in which the focus is on analyzing data on just one variable.

Multivariate Analysis
Any procedure in which the focus is on analyzing data on two or more variables simultaneously.

Bivariate Analysis
A special case of multivariate analysis involving just two variables.

Nonparametric Procedures
Analysis techniques and hypothesis tests that are appropriate when the data are nonmetric (i.e., nominal or ordinal).

Parametric Procedures
Analysis techniques and hypothesis tests that are appropriate only when the data are metric (i.e., interval or ratio).

The second factor affecting the choice of analysis techniques is the nature of the data collected. Particularly relevant in this regard is the measurement level of the data, that is, whether they are nominal, ordinal, interval, or ratio. As was discussed in Chapters 11 and 16, nominal and ordinal data (i.e., nonmetric data) are not as powerful or versatile as interval and ratio data (i.e., metric data). Therefore, only relatively crude statistical analyses can be performed with nonmetric data when compared with metric data.

The types of analyses and hypothesis tests appropriate for nonmetric data are typically labeled as ***nonparametric procedures***. Statistical procedures that are nonparametric require only minimal assumptions about the nature of the data, especially with respect to their measurement level and the shape of their distribution.

Analysis techniques suitable for metric data are said to be ***parametric procedures***. The use of most parametric methods requires data with at least interval-scale properties and having a distribution that resembles the normal probability distribution.[1]

In short, as a general rule, nonparametric procedures are appropriate for nominal and ordinal data, and parametric procedures are appropriate only for interval and ratio data. We will discuss several procedures of each type in this and the next two chapters.[2]

SPECIFIC HYPOTHESIS TESTS

This section deals with certain univariate tests that are used quite frequently. Table 17.2 presents an overview of the specific hypothesis tests

---

[1] Further discussion of the assumptions that must be met for use of parametric and nonparametric methods, as well as the pros and cons of the two methods, is available in several excellent books. See, for example, Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1956); Jean Dickinson Gibbons, *Nonparametric Methods for Quantitative Analysis* (New York: Holt, Rinehart and Winston, 1976).

[2] Selecting an appropriate analysis technique in a situation can be more complex than what our discussion here may suggest. However, a comprehensive presentation of all the intricacies involved in choosing a proper analysis procedure is very involved. An excellent discussion—one that gives a step-by-step flow diagram for selecting a technique after taking into account a variety of considerations—is given in Frank M. Andrews, Laura Klem, Terrence N. Davidson, Patrick M. O'Malley, and Willard L. Rodgers, *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data* (Ann Arbor, Mich.: Institute for Social Research, 1981).

Table 17.2 *Univariate Hypothesis Tests*

Type of Test	Primary Purpose of the Test	Illustrative Research Question the Test Is Designed to Answer
Nonparametric Tests		
Chi-square goodness-of-fit test	To see whether the observed distribution of responses on a *nominal-scaled* variable is significantly different from an expected distribution	Can the distribution of survey respondents with respect to various occupational categories be considered similar to the population distribution of occupational categories?
Kolmogorov-Smirnov test	To see whether the observed distribution of responses on an *ordinal-scaled* variable is significantly different from an expected distribution	Can the distribution of survey respondents with respect to different age categories be considered similar to the population distribution of age categories?
Parametric Tests		
Test for a single mean	To test hypotheses that compare the population mean of a variable to a prespecified value	Is the average waiting time for Ace Supermarket's customers at the store's checkouts significantly greater than 10 minutes?

(Continued)

we will discuss.[3] As shown by Table 17.2, two of the tests are nonparametric tests while the rest are parametric tests. We begin our discussion with the nonparametric tests.

Chi-square Goodness-of-Fit Test *A nonparametric test to check whether the observed distribution of sample data on a nominal-scaled variable is consistent with an expected distribution, that is, a distribution that would be expected if the sample had come from a prespecified population.*

Chi-square Goodness-of-Fit Test

The general purpose of the ***chi-square goodness-of-fit test*** is to see whether the distribution of data on a nominal-scaled variable is consistent with what one would expect if the sample had come from a prespecified population. The research question the test is capable of answering can be stated as follows: "Is the distribution implied by the sample data likely to be the same as that of some prespecified popula-

---

[3] The hypothesis tests listed in Table 17.2 are not meant to be exhaustive. Discussions of a greater variety of tests are given in many standard statistics books, such as Morris Hamburg, *Statistical Analysis* and Andrews, Klem, Davidson, O'Malley, and Rodgers, *Guide for Selecting Statistical Techniques for Decision Making* (New York: Harcourt Brace Jovanovich, 1977), pp. 271–275.

Table 17.2 *Continued*

Type of Test	Primary Purpose of the Test	Illustrative Research Question the Test Is Designed to Answer
Parametric Tests		
Test for a single proportion	To test hypotheses that compare the population proportion of a variable to a prespecified value	Is the proportion of households owning VCRs in Dallas significantly less than .3?
Test of two means	To test hypotheses that compare the population mean of a variable for two separate populations	Is the average expenditure per household on eating out significantly higher in Detroit than in Des Moines?
Test of two means when samples are dependent	To test hypotheses that compare two population means of the same variable when the data for the test are collected from the same set of sample units	On the basis of data collected from a panel of households before and after the special ad campaign for Suds detergent, is the mean purchase volume of Suds per household significantly higher after the campaign than before?
Test of two proportions	To test hypotheses that compare the population proportion of a variable for two separate populations	Is the proportion of two-income households significantly lower in the Midwest states than in the Northeast states?

tion?" A question like this one may arise under a variety of circumstances, as the following example illustrates.

> **Example:** A restaurant added four different fruit pies to its menu and monitored the unit sales of each over a period of time. Data on unit sales, broken out by type of pie, are now available. The restaurant's manager wants to know whether or not the sales data imply an equal customer preference for each type of pie. The underlying research question is whether the distribution of customer preferences implied by the sales data is consistent with one that might be expected from a population whose preferences are evenly divided across the four types of pies.

The chi-square goodness-of-fit test is appropriate in this example because the test involves comparing the actual sample distribution (or *observed distribution*, as it is usually called) with an *expected* sample distribution in order to see whether the two differ significantly. The expected sample distribution describes the manner in which the sample data would have been distributed if the sample had come from a prespecified population.

To illustrate, suppose these additional details about the previous example are available: The restaurant sold 600 fruit pies during one month. The observed distribution of sales was as follows:

Apple pies	180
Cherry pies	120
Peach pies	160
Blueberry pies	140
Total	600

The expected distribution in this case is derived by reallocating the total sales of 600 pies under the assumption of equal customer preference for all four varieties. The expected distribution of sales is as follows:

Apple pies	150
Cherry pies	150
Peach pies	150
Blueberry pies	150
Total	600

Let us now perform a chi-square goodness-of-fit test in the context of this example, assuming a significance level of .05 and following the seven-step procedure outlined earlier.

Step 1. State H_o and H_a. The null hypothesis can be stated in two different, but conceptually equivalent, forms:

H_o: The observed and expected distributions of pie sales are statistically equivalent.

Or

H_o: Customer preference for the four types of pies are identical.

The alternative hypothesis is

H_a: The observed and expected distributions of pie sales are not statistically equivalent.

Or

H_a: Customer preferences for the four types of pies are not identical.

Step 2. Identify the test statistic. Since the sample data are in nominal form, only a nonparametric test is appropriate for deciding whether to reject H_o or H_a. Specifically, the test statistic is the chi-square (χ^2) statistic, whose sampling distribution resembles the chi-square probability distribution.

Step 3. Determine whether the test is one-tailed or two-tailed. Hypothesis tests involving the chi-square statistic (at least the ones we will be dealing with) are *all* one-tailed. Thus the test in our example will be one-tailed.

Step 4. Determine the critical test statistic value. Two pieces of data are needed to determine the critical chi-square value (χ^2_c): The significance level (α, which is specified to be .05 in this case) and the so-called degrees of freedom (d.f.). The appropriate degrees of freedom in a univariate chi-square test are given by $k - 1$, where k is the number of nominal categories in the sample data distribution. Since there are four categories of pies in our example,

$$\text{d.f.} = 4 - 1 = 3.$$

Refer to the standard chi-square table in Appendix C. The critical chi-square value (χ^2_c) when $\alpha = .05$ and d.f. = 3 is 7.81.

Step 5. State the decision rule. The decision rule for chi-square tests is generally of the following form: "Reject H_0 if χ^2 (computed from sample data) is greater than χ^2_c." In our example, if the computed value of χ^2 exceeds 7.81, the restaurant manager should conclude that customer preferences for the four types of pies are not identical.

Step 6. Determine the test statistic value. The computed value of the test statistic in chi-square goodness-of-fit tests is derived by using the following expression:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i and E_i are the observed and expected *frequency counts*, respectively. Examining this expression carefully shows that the lower the discrepancies between the observed and the expected frequency counts, the lower the computed value of χ^2 will be. When the observed and the expected frequency distributions are identical, $\chi^2 = 0$. Therefore the smaller the computed value of χ^2, the weaker the evidence against H_o will be. Notice that this result is consistent with the general decision rule, that is, to reject H_o only when χ^2 is large enough to exceed χ^2_c.

The emphasis here on *frequency counts* is noteworthy. Chi-square values should always be computed by using the raw number of units (as opposed to percentages or proportions of units) falling in each category.

In our example we compute

$$\chi^2 = \frac{(180-150)^2}{150} + \frac{(120-150)^2}{150} + \frac{(160-150)^2}{150} + \frac{(140-150)^2}{150}$$

$$= 6.0 + 6.0 + .67 + .67 = 13.34.$$

Step 7. Reject H_o or H_a. Since $\chi^2(=13.34) > \chi^2_c(=7.81)$, the restaurant manager should reject H_o and conclude that customer preferences for the four types of pies are *not* identical. In fact, the computed χ^2 value (13.34) is so high that the probability of making a Type I error (i.e., the *actual* significance level) is less than .005 (verify this result by using Appendix C). Stated differently, if consumer preferences were indeed identical for the four types of pies, a chi-square value of 13.34 would have occurred less than five times in a thousand.

One final point about chi-square goodness-of-fit tests is in order. These tests are not meaningful when the *expected* frequency counts are too small. A rule of thumb is to ensure that each category has an expected frequency count of at least five units. Therefore, if the prespecified population is likely to be very sparse in certain categories and dense in others, the total sample size must be large enough to ensure an adequate expected sample frequency count in each category.

The rationale for the minimum expected frequency count requirement can be summarized as follows: Since the expected frequency count (E_i) appears in the denominator of the expression for χ^2, several categories with low expected frequency counts will artificially inflate the computed value of χ^2. This inflated value, in turn, may lead to the rejection of H_o without guaranteeing that the probability of a Type I error will be within α.

Kolmogorov-Smirnov Test

Kolmogorov-Smirnov Test
A nonparametric test to check whether the observed distribution of sample data on an ordinal-scaled variable is consistent with an expected distribution. This test is similar to the chi-suqare goodness-of-fit test, except that it requires ordinal data.

The ***Kolmogorov-Smirnov test*** is appropriate for examining the fit between observed and expected frequency distributions of data when the variable is ordinal-scaled. Consider the following situation.

Example. Suppose a random sample of 200 households drawn from a certain population has the following income distribution:

Less than $10,000	60
$10,000 to less than $20,000	50
$20,000 to less than $30,000	40
$30,000 to less than $40,000	30
$40,000 or more	20
Total	200

Also, suppose census data indicate that the percentages of population households in the five income categories (going from lowest to highest) are as follows: 35%, 30%, 15%, 10%, and 10%. On the basis of this data, is the sample representative of the population with respect to household income?

To answer the question posed in this example, we can, of course, perform a chi-square goodness-of-fit test by generating an expected frequency distribution of the sample of 200 households, using the census proportions, and comparing it with the observed frequency distribution. However, the chi-square goodness-of-fit test will not take into account that the variable categories are ordinal rather than nominal. In other words, the more refined quality of the available data will essentially be wasted if such a test is used.

Here is where the Kolmogorov-Smirnov test comes in. The Kolmogorov-Smirnov test is conceptually very similar to the chi-square goodness-of-fit test. *In both tests the hypothesis statements are identical, and the decision rules for rejecting H_o are of the same form.* However, there are some key differences in the computational procedures. Specifically, in a Kolmogorov-Smirnov test the observed and expected distributions are expressed as *proportions* (rather than frequency counts) and are converted to *cumulative* distributions before comparisons are made. These two steps are sumarized in Table 17.3 for our example.

The appropriate test statistic in this case is the Kolmogorov-Smirnov D-statistic, whose value is given by

$$D = \max_i |OCP_i - ECP_i|,$$

where D is the *maximum absolute discrepancy* (across all categories) between the observed cumulative proportion and the corresponding expected cumulative proportion. (From now on the specific steps in the hypothesis-testing procedure will not be explicity stated; however, you should be able to recognize them.)

Table 17.3 *Derivation of Observed and Expected Cumulative Proportion Distributions of Household Incomes*

Income Category (i)	Observed Frequency (O_i)	Observed Proportion (OP_i)	Expected Proportion* (EP_i)	Observed Cumulative Proportion (OCP_i)	Expected Cumulative Proportion (ECP_i)
1	60	0.30	0.35	0.30	0.35
2	50	0.25	0.30	0.55	0.65
3	40	0.20	0.15	0.75	0.80
4	30	0.15	0.10	0.90	0.90
5	20	0.10	0.10	1.00	1.00
	200	1.00	1.00		

*Based on census data.

For a significance level of .05, the critical value of the test statistic (D_c) can be approximated by the following expression, provided the sample size (n) is 36 or more:

$$D_c = \frac{1.36}{\sqrt{n}}.$$

When n is less than 36 and/or when α is different from .05, one can refer to standard tables that are available for determining the critical values of D for any given α and n.[4]

The decision rule in a Kolmogorov-Smirnov test is, "Reject H_o if $D > D_c$." In our example the maximum absolute discrepancy between the observed and expected cumulative proportions occurs for the second income category (see Table 17.3):

$$D = |.55 - .65| = .10.$$

Assuming $\alpha = .05$,

$$D_c = \frac{1.36}{\sqrt{200}} = \frac{1.36}{14.14} = .096.$$

Since $D > D_c$, H_o should be rejected: The sample chosen *cannot* be considered to be adequately representative of the population with respect to household income.

Unlike the chi-square goodness-of-fit test, the Kolmogorov-Smirnov test does not have any minimum expected frequency requirement for each category. Moreover, the computational procedure is somewhat less cumbersome. These additional advantages of the Kolmogorov-Smirnov test further favor its use when the variable categories are ordinal.[5]

When the raw data on variables of interest are stored on computer disks, the chi-square goodness-of-fit and Kolmogorov-Smirnov tests can be conducted by using a standard statistical package such as SPSS. In addition to the obvious computational ease, using the computer to conduct these tests will automatically yield the actual significance level or p-value for the test. Research in Use 17.1 presents and briefly describes several illustrative computer outputs pertaining to the two tests.

Test for a Single Mean

We have already examined in detail one instance where the focus was on hypotheses about a single mean. Specifically, in situation A described

---

[4] For further details, see, for example, Siegel, *Nonparametric Statistics.*

[5] For a more complete and theoretically rigorous discussion of the relative merits of the Kolmogorov-Smirnov test, see, for example, Gibbons, *Nonparametric Methods*, pp. 75–77. This reference also describes several other applications of the Kolmogorov-Smirnov test.

Research in Use 17.1 *Financial Services Company Study*

Chi-square Goodness-of-Fit and Kolmogorov-Smirnov Tests Using SPSS

A large financial services company conducted a survey of approximately 475 of its customers. The questionnaire contained several questions pertaining to customer demographics including sex and education.

The company's customer base consisted of roughly 50% males and 50% females. It wanted to verify whether the sample reflected this 50-50 split. A chi-square goodness-of-fit test of the sample data on sex (coded as 1 = Male and 2 = Female) was conducted using the SPSS computer program, and the following results were obtained.

```
                          CHI-SQUARE TEST
---------------------------------------------------------------
    SEX
                        CASES
    CATEGORY          OBSERVED      EXPECTED      RESIDUAL
           1               252        237.00         15.00
           2               222        237.00        -15.00
                           ---
        TOTAL              474
        CHI-SQUARE             D.F.            SIGNIFICANCE
             1.899                1                   0.168
---------------------------------------------------------------
```

The actual significance level (p-value) of .168 (under the label "SIGNIFICANCE") in the computer output implies that there is almost a 17% chance that the observed difference between the numbers of males and females in the sample could have occurred owing to chance even when the total customer base (i.e., the population) is evenly divided between males and females. Moreover, the null hypothesis that the sample adequately represents the population with respect to sex distribution will therefore not be rejected at the customary significance level of .05.

In this study education was coded as a four-category variable: 1 = high school or less; 2 = some college; 3 = college graduate; and 4 = graduate school. For a determination of whether the sample distribution with respect to education was significantly different from a uniform distribution (i.e., an expected proportion of .25 in each category), a chi-square goodness-of-fit test was conducted. The SPSS output for this test is shown next.

(Continued)

earlier, we tested the following hypotheses:

$$H_o\colon \mu \le \$12{,}000;$$
$$H_a\colon \mu > \$12{,}000.$$

Therefore we do not need to repeat the entire procedure here. Some clarification is needed, however, concerning the proper test statistic and critical value to use when conducting tests involving single means. Recall

Research in Use 17.1 *Continued*

```
                         CHI-SQUARE TEST

        ED

                          CASES
        CATEGORY        OBSERVED      EXPECTED      RESIDUAL
               1              89        117.50        -28.50
               2             126        117.50          8.50
               3             151        117.50         33.50
               4             104        117.50        -13.50
                             ---
           TOTAL             470
          CHI-SQUARE              D.F.          SIGNIFICANCE
             18.630                 3                  0.000
```

The actual significance level of .000 implies that the odds are less than 1 in 1000 that the difference between the observed distribution and a uniform distribution could have occurred owing to chance.

You probably noticed that education is an *ordinal-scaled* variable. Therefore the Kolmogorov-Smirnov test is more appropriate than the chi-square goodness-of-fit test. The SPSS output for the Kolmogorov-Smirnov test is shown next.

```
             KOLMOGOROV—SMIRNOV GOODNESS-OF-FIT TEST

    ED
    TEST DISTRIBUTION—UNIFORM               RANGE: 1 TO 4
                               CASES: 470
                        MOST EXTREME DIFFERENCES
    ABSOLUTE      POSITIVE      NEGATIVE       K-S Z       2-TAILED P
     0.22128       0.13936      -0.22128       4.797            0.000
```

The actual significance level of .000 (under the label "2-TAILED P") leads us to the same conclusion as before. Since the Kolmogorov-Smirnov test is a more refined test than the chi-square goodness-of-fit test, it will invariably detect a difference between the distributions whenever the chi-square goodness-of-fit does so. However, the reverse is not necessarily true—i.e., the chi-square goodness-of-fit test, because it is less refined, may not always detect a difference whenever a Kolmogorov-Smirnov test does so. For this reason it is preferable to use the Kolmogorov-Smirnov test when the variable categories are ordinal.

that in situation A the test statistic we used was the *z*-statistic (standard normal deviate). Our use of the *z*-statistic was actually an approximation, because the theoretically correct sampling distribution for sample means is the *t-distribution*, and the appropriate test statistic is the *t-statistic*. Fortunately, when the sample size is 30 or more the *t*-distribution closely resembles the standard normal curve. Hence our use of the z-statistic in situation A was justified.

If the sample size in situation A was, say, only 25 instead of 400, the z-statistic would no longer be appropriate. The correct test statistic is the t-statistic, defined as

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}.$$

Also, the critical test statistic value (t_c) must be obtained from a t-table (Appendix D at the end of this book). The value of t_c depends not only on α but also on the number of degrees of freedom. The t-statistic associated with a single sample mean has $n - 1$ degrees of freedom, where n is the sample size.

To illustrate the use of the t-statistic, let us rework some of the steps in the problem for situation A, assuming $n = 25$, $\bar{x} = \$13,000$ (same as before), and $s = \$8000$ (same as before). From the t-table in Appendix D, $t_c = 1.71$ for $\alpha = .05$ and d.f. $= 24$. Hence the decision rule is, "Reject H_o if $t > 1.71$." Let us now compute the value of t from the sample data:

$$t = \frac{13,000 - 12,000}{8000/\sqrt{25}} = \frac{1000}{1600} = .625.$$

Since the computed value of t is less than 1.71, H_o *cannot* be rejected. In other words, Karen should not introduce the product line into the new market area.

Notice that this conclusion contradicts the one we arrived at earlier after performing a z-test in the same situation (except for the larger sample size). However, bear in mind that by failing to reject H_o based on the t-test we may be committing a Type II error. Indeed, in general, when everything else remains the same, the smaller the sample size, the less likely it is that H_o will be rejected, the higher will be the β-probability, and the lower will be the power. Since traditional hypothesis-testing procedures only limit the probability of committing a Type I error, one has to be cautious in interpreting a finding that H_o cannot be rejected on the basis of a small-sample-size hypothesis test.

Test for a Single Proportion

Consider the following example.

Example. Ms. Jones is marketing vice president for Peripherals, Inc., a firm that sells a variety of personal-computer accessories under the brand name Comp-Ease. She is wondering whether she should yield to a request from Mr. Berry, the advertising manager, for a substantial increase in the firm's advertising budget. Mr. Berry had just surveyed a random sample of 100 personal-computer owners and found that only 20 of them had heard of the Comp-Ease name. In making his plea for an increased advertising budget, Mr. Berry had described the sample awareness rate of .2 as being "shamefully low." Ms. Jones, however, feels that the .2 awareness rate is not all that bad, especially since the Comp-Ease line had been introduced only recently. In any case, she is reluctant to increase the advertising budget unless Mr. Berry can show "beyond a reasonable doubt" (i.e., with less than a 5% chance of error) that the true awareness rate for the Comp-Ease name across all personal-computer owners is less than .3. Should Ms. Jones increase the advertising budget on the basis of Mr. Berry's survey results?

The key decision in this scenario boils down to testing the following hypotheses about the population proportion (π is the symbol for population proportion) of personal-computer owners who are aware of the Comp-Ease name:

$$H_o: \pi \geq .3;$$

$$H_a: \pi < .3.$$

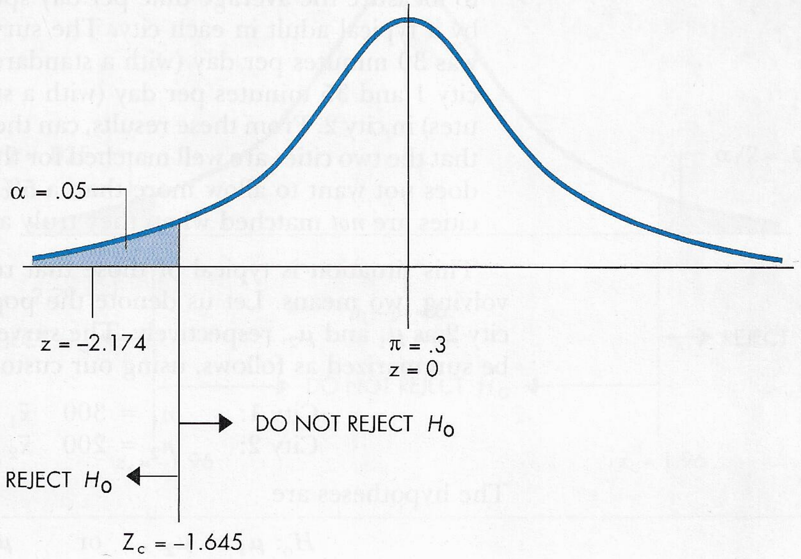
The theoretically correct sampling distribution for sample proportions is the binomial distribution. However, for a sufficiently large sample size the binomial distribution resembles the normal distribution, and the z-statistic can be used as the test statistic. How large a sample size is "sufficiently large"? A rule of thumb is that n must be large enough to make each of the following quantities equal at least 10: $n\pi$ and $n(1 - \pi)$, where the value of π is its value when H_o is just true. In our scenario, $n = 100$ and $\pi = .3$. Hence $n\pi$ and $n(1 - \pi)$ are both greater than 10, and the z-statistic, as defined below, can be used as the test statistic:

$$z = \frac{p - \pi}{\sqrt{\pi(1 - \pi)/n}},$$

where p is the sample proportion. Notice that the denominator of this expression is really the standard error of the sample proportion. However, in contrast to the standard error in the test for the mean, where we had to approximate $\sigma_{\bar{x}}$ by $s_{\bar{x}}$ (which, of course, can only be determined after the sample standard deviation is computed), the standard error of the proportion is known exactly once π and n are specified.

Since only small values of p relative to π (or relatively large *negative* values of z) will lead to the rejection of the H_o stated earlier, this hypothesis test is one-tailed, as shown in Exhibit 17.4. The critical test statistic value (z_c) is negative because the shaded tail corresponding to rejection of H_o is on the left-hand side of the sampling distribution centered on $\pi = .3$, or $z = 0$ (the z-value at the center of a normal probability distribution is always zero; z-values on the left-hand side of the distribution are all negative, while those on the right-hand side are all positive). From Appendix B, $z_c = -1.645$ for $\alpha = .05$. The appropriate decision rule here is, "Reject H_o if $z < -1.645$." Using $p = .2$,

Exhibit 17.4 *Hypothesis Test Related to Proportion of Personal-Computer Owners*



$\pi = .3$, and $n = 100$, we can compute the value of z:

$$z = \frac{.2 - .3}{\sqrt{(.3)(.7)/100}} = \frac{-.1}{.046} = -2.174.$$

Since $-2.174 < -1.645$, we reject H_o; The sample awareness rate of .2 is too low to support the hypothesis that the population awareness rate is .3 or more. The actual significance level (p-value) corresponding to $z = -2.174$ is approximately .015 (from Appendix B). This level of significance implies that the odds are less than 15 in 1000 that the sample awareness rate of .2 would have occurred entirely owing to chance (i.e., when the population awareness rate is .3 or higher). In other words, Mr. Berry apparently has a legitimate request, and Ms. Jones should increase the advertising budget.

Test of Two Means

Consider the situation described in the following example.

Example. A health service agency has designed a public service campaign to promote physical fitness and the importance of regular exercise. Since the campaign is a major one, the agency wants to make sure of its potential effectiveness before running it on a national scale. To conduct a controlled test of the campaign's effectiveness, the agency has identified two similar cities: City 1 will serve as the test city, and city 2 will serve as a control city. A preliminary random survey of 300 adults in city 1 and 200 adults in city 2 was conducted to measure the average time per day spent on some form of exercise by a typical adult in each city. The survey showed that this average was 30 minutes per day (with a standard deviation of 22 minutes) in city 1 and 35 minutes per day (with a standard deviation of 25 minutes) in city 2. From these results, can the agency conclude confidently that the two cities are well matched for the controlled test? The agency does not want to allow more than a 5% chance of inferring that the cities are *not* matched when they truly are matched.

This situation is typical of those that require testing hypotheses involving two means. Let us denote the population means in city 1 and city 2 as μ_1 and μ_2, respectively. The survey results in the two cities can be summarized as follows, using our customary notation:

City 1: $n_1 = 300$ $\bar{x}_1 = 30$ $s_1 = 22$
City 2: $n_2 = 200$ $\bar{x}_2 = 35$ $s_2 = 25$

The hypotheses are

$$H_o\colon \mu_1 = \mu_2 \quad \text{or} \quad \mu_1 - \mu_2 = 0;$$
$$H_a\colon \mu_1 \neq \mu_2 \quad \text{or} \quad \mu_1 - \mu_2 \neq 0.$$

In a two-means hypothesis test, when the sample sizes are sufficiently large (i.e., $n_1 > 30$ and $n_2 > 30$), the sampling distribution for the difference between sample means resembles the normal probability distribution, and the test statistic is the *z*-statistic, given by

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}.$$

Both n_1 and n_2 are greater than 30 in our example. The *z*-statistic can therefore be used as the test statistic.

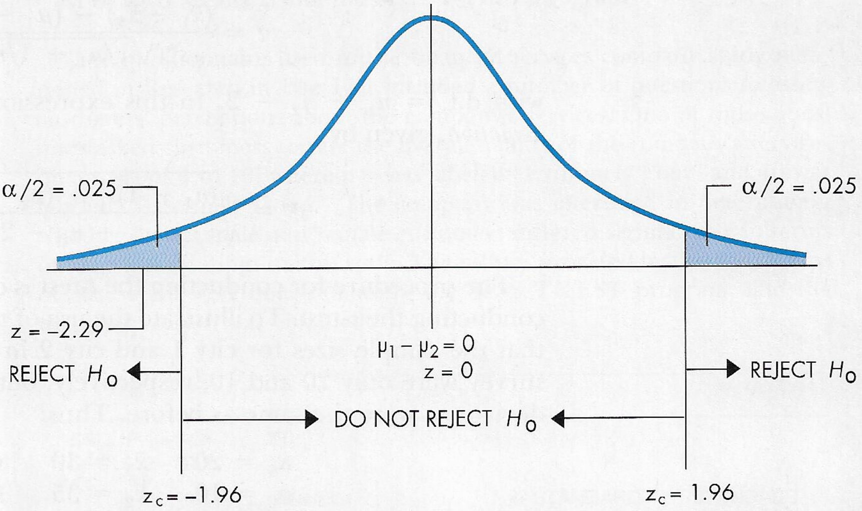
Since the null hypothesis is in the form of a strict equality, the hypothesis test is two-tailed, and hence we have to identify two critical values of *z*, one for each tail of the sampling distribution. The probability corresponding to each tail is .025, since $\alpha = .05$. From Appendix B the *z*-value associated with a tail probability of .025 is 1.96. The decision rule is, "Reject H_o if $z < -1.96$ or if $z > 1.96$."

Let us now compute the value of *z* from the survey results and under the customary assumption that the null hypothesis is true (i.e., $\mu_1 - \mu_2 = 0$):

$$z = \frac{(30 - 35) - (0)}{\sqrt{(22)^2/300 + (25)^2/200}} = \frac{-5}{\sqrt{1.61 + 3.13}}$$
$$= \frac{-5}{\sqrt{4.74}} = \frac{-5}{2.18} = -2.29.$$

Since $z < -1.96$, we should reject H_o. Therefore the health service

Exhibit 17.5 *Hypothesis Test Related to Mean Exercising in Two Cities*



agency cannot confidently assume that the current levels of exercising activity of residents in the two cities are identical. A pictorial summary of this hypothesis test is shown in Exhibit 17.5.

In addition to the assumption of a large sample size, the use of the z-statistic in a hypothesis test involving two means requires the assumption of *independent samples*; that is, the samples must be chosen independently of each other. Thus, for example, the z-test is not appropriate for checking whether there is a statistically significant difference between two mean values obtained *at different times* from the *same* sample. Under such circumstances a different hypothesis-testing procedure, which we will discuss a little later, should be used.

Furthermore, just as in the case of a single-mean hypothesis test, when one or both sample sizes are small (i.e., $n_1 < 30$ or $n_2 < 30$), a t-test rather than a z-test should be conducted. Using the t-statistic in a two-means hypothesis test requires two additional assumptions, however:

1. The two populations from which the samples are selected are each normally distributed with respect to the variable.
2. The two populations have equal variances.

These assumptions are usually taken for granted. Ideally, however, they should be verified to ensure that conducting a t-test is legitimate.[6]

---

[6] Procedures for verifying the assumptions from sample data can be found in basic statistics textbooks, such as Neil Weiss and Matthew Hassett, *Introductory Statistics* (Reading, Mass.: Addison-Wesley, 1982).

When the two above assumptions, along with the assumption of independent samples, are satisfied, the appropriate test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s^*(\sqrt{1/n_1 + 1/n_2})},$$

with d.f. $= n_1 + n_2 - 2$. In this expression s^* is the *pooled standard deviation*, given by

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

The procedure for conducting the t-test is quite similar to the one for conducting the z-test. To illustrate the use of the t-statistic, let us assume that the sample sizes for city 1 and city 2 in the health service agency survey were only 20 and 10, respectively, but that means and standard deviations were the same as before. Thus:

$$\begin{array}{lll} n_1 = 20 & \bar{x}_1 = 30 & s_1 = 22 \\ n_2 = 10 & \bar{x}_2 = 35 & s_2 = 25 \end{array}$$

The degrees of freedom for the t-statistic are

$$\text{d.f.} = n_1 + n_2 - 2 = 20 + 10 - 2 = 28.$$

From Appendix D the critical value of t with 28 degrees of freedom for a tail probability of .025 is 2.05. Therefore the decision rule is, "Reject H_o if $t < -2.05$ *or* if $t > 2.05$." The pooled standard deviation is

$$\begin{aligned} s^* &= \sqrt{\frac{(20 - 1)(22)^2 + (10 - 1)(25)^2}{20 + 10 - 2}} \\ &= \sqrt{\frac{9196 + 5625}{28}} = \sqrt{\frac{14821}{28}} \\ &= \sqrt{529} \text{ (approximately)} = 23. \end{aligned}$$

So the test statistic is

$$\begin{aligned} t &= \frac{(30 - 35) - (0)}{23(\sqrt{\frac{1}{20} + \frac{1}{10}})} = \frac{-5}{23(\sqrt{.05 + .1})} \\ &= \frac{-5}{23(.39)} = \frac{-5}{8.97} = -.56. \end{aligned}$$

Since t is neither less than -2.05 nor greater than 2.05, we cannot reject H_o. In other words, the sample evidence is not strong enough to conclude that the two cities differ in terms of levels of exercising activity of their residents. As was pointed out under t-tests for single means, however, this inference must be viewed with a good deal of caution because the small sample sizes greatly increase the possibility of a Type II error.

Research in Use 17.2 illustrates a computer-conducted test of the

Research in Use 17.2 *Financial Services Company Study*

Test of Two Means Using the SPSS T-TEST Program

The questionnaire used in the financial services company study mentioned in Research in Use 17.1 included a number of questions to assess customers' perceptions about the company's services. One of those questions asked customers to rate the overall quality of the company's services on a scale of 1 to 10, wherein 1 was labeled "Extremely Poor" and 10 was labeled "Extremely Good." The company was interested in determining whether or not male and female customers differed significantly in terms of their mean ratings on this scale. The ratings provided by the two groups of customers were analyzed using the SPSS T-TEST program, and the following output was obtained.

```
----------------------------------------------------------------------
                               T-TEST
----------------------------------------------------------------------
   GROUP 1—SEX      EQ      1.
   GROUP 2—SEX      EQ      2.

                  NUMBER                   STANDARD        STANDARD
 VARIABLE       OF CASES       MEAN        DEVIATION          ERROR
 QUALITY
    GROUP 1          201     7.2886           1.946           0.137
    GROUP 2          157     7.1783           1.900           0.152

                 |    POOLED VARIANCE        |    SEPARATE VARIANCE
                 |       ESTIMATE            |       ESTIMATE
                 |        DEGREES            |        DEGREES
   F      2-TAIL |   T        OF     2-TAIL  |   T        OF      2-TAIL
 VALUE    PROB.  | VALUE   FREEDOM   PROB.   | VALUE   FREEDOM    PROB.
  1.05    0.756  |  0.54       356   0.591   |  0.54    338.91    0.590
----------------------------------------------------------------------
```

On the 10-point scale, males (GROUP 1) gave a mean rating of approximately 7.29, while females (GROUP 2) gave a mean rating of approximately 7.18. The lower half of the output contains three separate sections, each with an associated actual significance level or *p*-value (under the label "2-TAIL PROB."). The first section is a summary of a test (which has an *F*-value as the test statistic) conducted by the program to see if the variances for the two groups can be assumed to be equal. The *p*-value of .756 indicates that the null hypothesis of equal variances for the two groups cannot be rejected at the customary significance level of .05. In fact, the *p*-value of .756 implies that the odds are greater than 75 to 100 that any observed difference (which happens to be quite small, as reflected by the similar standard deviations shown in the upper half of the output) occurred entirely from chance. Therefore computing a *pooled* estimate of the standard deviation (the s^* value defined in the chapter) is appropriate.

(Continued)

Research in Use 17.2 *Continued*

The *t*-value based on such an estimate and the associated degrees of freedom and *p*-value are reported in the second section of the lower half of the output. The *p*-value implies that the odds are 59 to 100 that a difference of magnitude .11 (i.e., 7.29 − 7.18) could have occurred from chance. Therefore the null hypothesis of no difference between the two groups cannot be rejected at the customary significance level of .05.

The third section of the lower half of the output is similar to the second section except for the fact that the program computed the standard deviation used in figuring the *t*-value by using a formula that does *not* assume equality of variances for the two groups. This section (rather than the second section) should be examined if the *F*-value (for testing the equality of variances) turns out to be statistically significant (i.e., the null hypothesis of equality of variances is rejected).

difference between two independent means using the SPSS T-TEST program. The T-TEST program in SPSS also has an option for using the hypothesis-testing procedure to be discussed next.

Test of Two Means When Samples Are Dependent

The testing procedures in the preceding section can be used only when the two samples do not overlap and are chosen independently of each other. However, marketers often need to check for significant differences between two mean values when the samples are not independent. For instance, "On the basis of a survey of husband-wife households, is there a significant difference between the mean attitude score of husbands and that of wives toward our product?" Here, although there is a sample of husbands and a sample of wives, the two samples are not independent because they are from the *same sample of households*. Marketers also may want to check differences between two mean values when the two sets of data are from the *same sample of respondents* (i.e., the same individuals). For example, "On the basis of data from a panel of women, is the mean frequency of use for our product among the women significantly higher after the coupon campaign than before?" In both cases a modified hypothesis-testing procedure is necessary, as discussed next. The following example illustrates yet another situation in which the modified procedure is needed.

Example. A retail chain ran a special promotion in a representative sample of 10 of its stores, hoping to boost sales substantially. Weekly sales per store before and after the introduction of the special promotion are shown in Table 17.4. The key question now is whether the sample evidence in Table 17.4 is strong enough to support man-

Table 17.4 *Sales Per Store Before and After a Promotional Campaign*

Store Number (i)	Sales Per Store (× \$1000) Before Promotion (X_{bi})	Sales Per Store (× \$1000) After Promotion (X_{ai})	Change in Sales (× \$1000) ($X_{di} = X_{ai} - X_{bi}$)
1	250	260	10
2	235	240	5
3	150	151	1
4	145	140	−5
5	120	124	4
6	98	100	2
7	75	70	−5
8	85	95	10
9	180	200	20
10	212	220	8
Total			50

agement's a priori hunch that the special promotion would lead to a significant increase in sales.

For this example the "before" and "after" sales data (X_{bi} and X_{ai}, respectively) are clearly not independent since they are from the same sample of stores. The hypothesis-testing procedure in situations like these calls for first computing *difference* scores for pairs of related sample data; in our example these differences are the X_{di} values in the last column of Table 17.4. This step essentially collapses what appear to be two-sample data into a set of single-sample difference scores. The rest of the procedure is similar to the hypothesis-testing procedure for a single mean.

Let μ_d represent the population *mean change* in sales per store. Since management had a prior expectation that the special promotion would be effective, that expectation should be consistent with the alternative hypothesis (the null hypothesis, which should be stated more conservatively, is that the special promotion did not improve sales). Therefore:

$$H_o\colon \mu_d \le 0;$$
$$H_a\colon \mu_d > 0.$$

The sample estimate of μ_d is $\bar{x}_d$, given by

$$\bar{x}_d = \frac{\sum_{i=1}^{n} X_{di}}{n},$$

where n is the sample size. In our example, the summation in the numerator is the total of the difference scores in the last column in Table 17.4 (i.e., 50) and $n = 10$. Therefore $\bar{x}_d = 5$.

The sampling distribution for $\bar{x}_d$ is the t-distribution, and the test statistic is

$$t = \frac{\bar{x}_d - \mu_d}{s/\sqrt{n}},$$

with $(n - 1)$ degrees of freedom, and where s is the standard deviation of the difference scores and is given by

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_{di} - \bar{x}_d)^2}{n - 1}}.$$

Given the way H_o and H_a are stated, only relatively high values of $\bar{x}_d$ (or t) will lead to the rejection of H_o. Hence the hypothesis test is one-tailed, as shown in Exhibit 17.6, in which a significance level of .05 has been assumed. From Appendix D the critical value of t with $n - 1 = 9$ degrees of freedom, for a one-tail probability of .05, is 1.83. The decision rule then is, "Reject H_o if $t > 1.83$."

The standard deviation of the difference scores (s) in our example turns out to be 7.53. We can now compute the value of the test statistic:

$$t = \frac{\bar{x}_d - \mu_d}{s/\sqrt{n}} = \frac{5 - 0}{7.53/\sqrt{10}}$$
$$= \frac{5}{2.38} = 2.10.$$

Since $t > 1.83$, we reject H_o and conclude that the mean change in sales per store was significantly greater than zero. In other words, the special promotion was indeed effective.

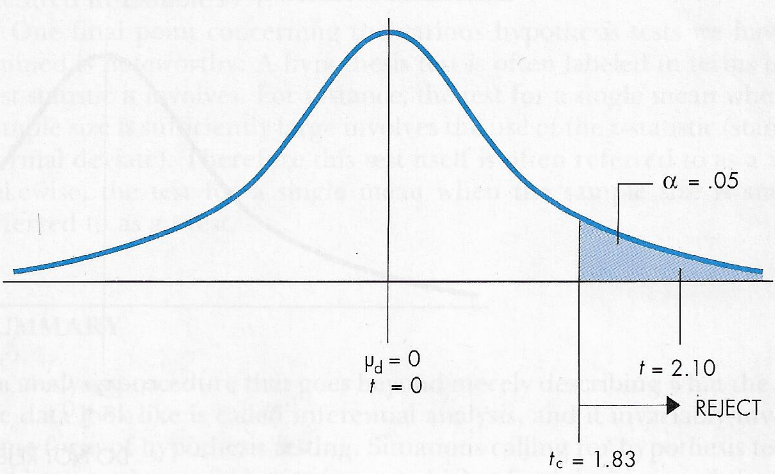
Test of Two Proportions

We will use the data in situation B (described earlier in the chapter) to illustrate hypothesis tests involving two proportions. In situation B, recall that Tom, the advertising manager for a frozen-foods company, is wondering whether commercial Y will be more effective than commercial X. The commercials were run in two comparable test cities. Subsequent surveys in the two cities showed that the sample awareness rates for commercials Y and X were 25% and 20%, respectively.

Using the subscript 1 for symbols pertaining to commercial Y and the subscript 2 for symbols pertaining to commercial X, we can summarize situation B as follows:

Sample sizes:	$n_1 = 200$	$n_2 = 200$
Sample proportions:	$p_1 = .25$	$p_2 = .20$

Exhibit 17.6 *Hypothesis Test Related to Change in Weekly Sales Per Store*



The hypotheses are

$$H_o\colon \pi_1 \le \pi_2 \quad \text{or} \quad \pi_1 - \pi_2 \le 0;$$
$$H_a\colon \pi_1 > \pi_2 \quad \text{or} \quad \pi_1 - \pi_2 > 0.$$

When the two samples are sufficiently large in a test involving two proportions, the sampling distribution of the difference between sample proportions can be assumed to be the normal probability distribution, and the *z*-test is appropriate. A rule of thumb for ensuring that the sample sizes are adequate is to verify that each of the following quantities is at least 10: n_1p_1, $n_1(1 - p_1)$, n_2p_2, and $n_2(1 - p_2)$. The data in situation B satisfy these requirements. Hence the *z*-test is appropriate. The test statistic is given by

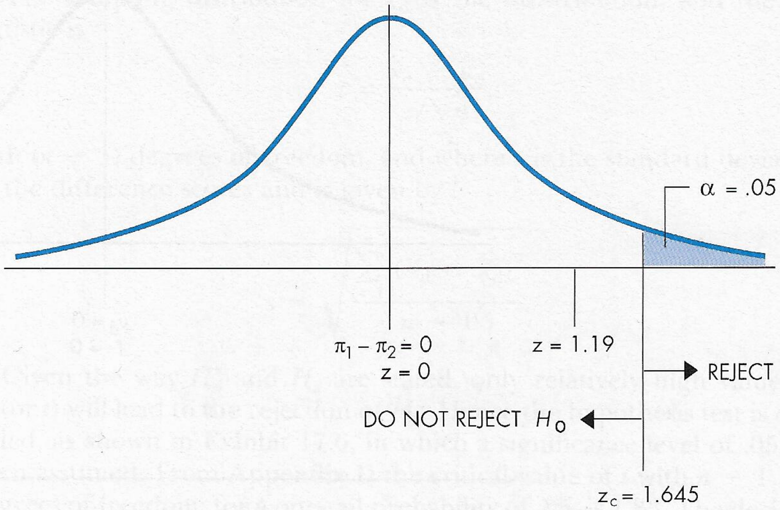
$$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sigma_{p_1 - p_2}},$$

where $\sigma_{p_1-p_2}$ is the population standard error for the difference between proportions. Since $\sigma_{p_1-p_2}$ is invariably unknown, it is estimated by the sample standard error ($s_{p_1-p_2}$), as follows:

$$s_{p_1-p_2} = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

where *P* is the weighted proportion across both samples and *Q* is its

Exhibit 17.7 *Hypothesis Test Related to Awareness Generated by Two Commercials*



complement:

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$Q = 1 - P$$

The null hypothesis in situation B will be rejected only for certain positive values of the quantity $p_1 - p_2$. Therefore the hypothesis test is one-tailed. Assuming the customary significance level of .05, the critical value of z (from Appendix B) is 1.645. The decision rule is, "Reject H_o if $z > 1.645$."

To compute the value of z from the sample data, we first have to compute P and Q, then $s_{p_1 - p_2}$ and z:

$$P = \frac{200(.25) + 200(.2)}{200 + 200} = \frac{50 + 40}{400} = \frac{90}{400} = .225$$

$$Q = 1 - .225 = .775$$

$$s_{p_1 - p_2} = \sqrt{(.225)(.775)(\tfrac{1}{200} + \tfrac{1}{200})}$$

$$= \sqrt{(.225)(.775)(\tfrac{1}{100})} = \sqrt{.00174} = 0.42$$

$$z = \frac{(.25 - .20) - (0)}{.042} = \frac{.05}{.042} = 1.19$$

Since $z < 1.645$, we *cannot* reject H_o. Therefore the sample evidence is not strong enough to suggest that commercial Y will be more effective than commercial X. The results of our hypothesis test in situation B are pictured in Exhibit 17.7.

One final point concerning the various hypothesis tests we have examined is noteworthy: A hypothesis test is often labeled in terms of the test statistic it involves. For instance, the test for a single mean when the sample size is sufficiently large involves the use of the *z*-statistic (standard normal deviate). Therefore this test itself is often referred to as a *z*-test. Likewise, the test for a single mean when the sample size is small is referred to as a *t*-test.

SUMMARY

An analysis procedure that goes beyond merely describing what the sample data look like is called inferential analysis, and it invariably involves some form of hypothesis testing. Situations calling for hypothesis testing will have a prespecified criterion on which a final decision, or inference, hinges. This criterion plays a key role in the formal statement of the null and alternative hypotheses.

Two kinds of errors are possible in testing hypotheses about population characteristics based on sample data: A Type I error, which refers to rejecting the null hypothesis when it is actually true, and a Type II error, *which refers to failing to reject the null hypothesis when it is false.* Traditional hypothesis-testing procedures place an upper bound on the probability of a Type I error—called the significance level α—but do not explicitly control Type II error. When α is lowered, however, the probability of a Type II error (called β) will increase and power $(1 - \beta)$, which represents protection against a Type II error, will decrease. The value of β depends not only on α but also on how false the null hypothesis really is.

The general procedure for testing hypotheses consists of the following steps:

1. Set up H_o and H_a (in accordance with the guidelines stated in the chapter).

2. Identify the nature of the sampling distribution of the variable and specify the appropriate test statistic.

3. Determine whether the hypothesis test is one-tailed or two-tailed (in general, a test will be one-tailed when H_o does not involve a strict equality).

4. Specify the critical value(s) for the test statistic, depending on the preset significance level.

5. State the decision rule for rejecting H_o.

6. Compute the value of the test statistic from sample data.

7. Compare the test statistic value with the critical value(s) and decide whether or not to reject H_0.

Hypothesis testing is an integral, although sometimes an implicit, part of virtually all statistical analyses that go beyond mere data description. Therefore factors that affect the choice of an appropriate analysis technique are also germane to selecting suitable hypothesis tests. Two such factors are crucial: number of variables and measurement level of the data. Analyses and hypothesis tests pertaining to just one variable are univariate; those pertaining to two or more variables are multivariate (they are sometimes described as bivariate when just two variables are involved). Nonparametric tests are used when the data are nonmetric (nominal or ordinal), and parametric tests are used when the data are metric (interval or ratio).

In this chapter we discussed several univariate hypothesis tests; the first two were nonparametric and the rest were parametric. The specific tests covered were as follows:

1. The chi-square goodness-of-fit test for comparing the observed and expected frequency distributions of a nominal-scaled variable.
2. The Kolmogorov-Smirnov test for comparing the observed and expected frequency distributions of an ordinal-scaled variable.
3. The z-test for a single mean when the sample size is large.
4. The t-test for a single mean when the sample size is small.
5. The z-test for a single proportion (valid only when the sample size is large).
6. The z-test for two means when both samples are large.
7. The t-test for two means when either sample is small.
8. The t-test for two means when the samples are dependent.
9. The z-test for two proportions (valid only when both samples are large).

The risk of committing a Type II error is especially high in the t-tests involving small sample sizes.

KEY TERMS

Descriptive analysis
Inferential analysis
Type I error
Type II error
Significance level
Power
Decision rule
Test statistic
One-tailed hypothesis test
Directional hypothesis
Nondirectional hypothesis
Two-tailed hypothesis test
Univariate analysis
Multivariate analysis
Bivariate analysis
Nonparametric procedures
Parametric procedures
Chi-square goodness-of-fit test
Kolmogorov-Smirnov test

QUESTIONS

1. Can data gathered through a descriptive-research project be useful in conducting inferential analysis? Why or why not? (Refer to Chapter 4 and refresh your memory about what is meant by descriptive research.)
2. What are Type I and Type II errors? Explain the relationship between these two types of errors.
3. What will happen to the probability of committing a Type II error when the significance level is increased from, say, .05 to .1?
4. "Hypothesis-testing procedures, given their statistical rigor, are free of decision maker subjectivity." Discuss this statement.
5. A firm wanted to test four alternative types of packaging—A, B, C, and D—for one of its products. Specifically, it wanted to determine whether there was any significant difference in consumer preference across the four types of packaging. It conducted a consumer preference test in which a sample of 200 consumers participated. Each participant was shown the four types and asked to pick the one that he or she preferred the most. The results of this test were as follows: Of the 200 consumers, 45 preferred A, 55 preferred B, 55 preferred C, and 45 preferred D. From these results, can one conclude with 99% confidence that there is some significant difference in preferences across the four types of packaging?
6. The representative sample selected for a mail survey of businesses had the following distribution of types of firms: 15% industrial-goods manufacturers, 35% consumer goods manufacturers, 25% retailing firms, and 25% service firms. A total of 1000 questionnaires were mailed, but responses were obtained only from 280 firms, consisting of 60 industrial-goods manufacturers, 80 consumer-goods manufacturers, 55 retailing firms, and 85 service firms. Assuming that the type of firm is likely to have a major bearing on the questionnaire responses, can one be 95% confident that there was no significant *nonresponse error*? (For a definition of nonresponse error, see Chapter 15.)
7. From census information the households in a certain city have the following distribution with respect to the number of people in them:

Category	Number of People	Percentage of Households
A	1	10%
B	2	30%
C	3	30%
D	4	20%
E	More than 4	10%

A sample of 500 households chosen from the city was distributed across the five categories as follows: A = 35, B = 110, C = 160, D = 110, E = 85. Can this sample be considered statistically equivalent to the population with respect to household size (at a significance level of .05)?

8. The Miller Cereal Company has been running an advertising campaign for one of its brands for quite some time. It wants to continue running the campaign only if at least 20% of the target audience is now aware of the brand; otherwise, it wants to discontinue the campaign. A just-completed survey of a random sample of 500 customers in the target audience showed that 90 of them were aware of the advertised brand. Should the Miller Cereal Company discontinue the advertising campaign on the basis of this evidence? Assume that if the company does decide to discontinue the campaign, it wants to be 95% confident that it is making the correct decision.

9. Questions 6a and 6b in Chapter 12 are problems dealing with estimating confidence intervals. Rework those problems by using the hypothesis-testing framework discussed in this chapter. Are the conclusions stemming from the hypothesis-testing procedure the same as those stemming from the confidence intervals estimated earlier? Why or why not?

10. The XYZ Company wants to advertise its product in a magazine only if the average age of the magazine's readership is greater than 40. A random sample of 625 readers of the magazine had an average age of 41, with a standard deviation of 20. Suppose the firm does not want more than a 5% chance of making the mistake of advertising in the magazine when the magazine is really inappropriate. Set up, and test, a suitable pair of hypotheses, and indicate whether the firm should advertise in the magazine.

11. ABC Company is currently selling brand Z through 3000 retail outlets across the United States. To strengthen Z's market position, ABC's objective is to make it the highest-selling brand in each of the 3000 outlets. To achieve this goal, one of ABC's options is to cut the retail selling price of Z. However, it wants to cut the retail price only if analysis of last year's sales indicates that Z is *not* the highest-selling brand in more than 50% of the stores. In its decision making based on this sample information, ABC does not want more than 5% chance of cutting price when Z is indeed the highest-selling brand in at least 50% of the 3000 stores. Suppose ABC picks a random sample of 400 stores and finds that Z is not the highest-selling brand in 210 of them. Should the price be cut based on this sample information?

12. A firm wanted to see if there were a significant difference between two neighboring communities in terms of the mean number of members per household. A random survey of 100 households in the first community showed that the mean household size was 3 with a standard deviation of 5. In the second community a random survey of 90 households showed that the mean household size was 4 with a standard deviation of 6. Assuming a significance level of .01, test the hypothesis that the two communities are statistically equivalent with respect to household size.

13. A sample of 400 female consumers of a product had a mean age of 33 with a standard deviation of 5 and a sample of 500 male consumers had a mean age of 32 with a standard deviation of 3. Can the mean ages of the female and male populations of consumers be considered significantly different at a significance level of .05?

14. A new product was promoted to two similar target markets, using two *different* advertising media: Medium 1 was used for market 1, and medium 2 was used for market 2. After the promotional campaign was over, a random sample of 200 people was surveyed from market 1, and a random sample of 300 people was surveyed from market 2. Of the market 1 sample 154 were aware of the new product, while of the market 2 sample 210 were aware of the new product. Using this evidence, test the *null* hypothesis that the effectiveness of medium 1 is less than or equal to that of medium 2 in terms of generating awareness among consumers (at the .05 significance level).

Relationship of Power to Significance Level and Population Parameter Value

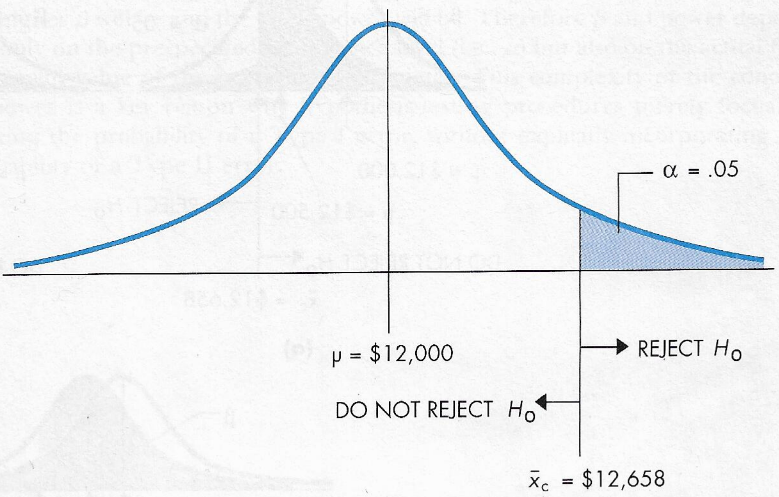
The hypothesis-testing procedures discussed in the chapter focused exclusively on limiting the probability of a Type I error. This emphasis does not mean that Type II errors are unimportant. The risk of committing a Type II error (i.e., failing to reject H_o when it is false) is very real and is present in all hypothesis tests. Indeed, as will be shown in this appendix, lowering the probability of a Type I error will increase the probability of a Type II error.[1] Researchers invariably constrain the probability of a Type I error to be within a small, prespecified significance level and let the probability of a Type II error fall where it may. This point is noteworthy because it highlights a bias inherent in the traditional hypothesis-testing procedure—namely, it reflects the posture of giving the null hypothesis the benefit of the doubt and rejecting it only when the sample evidence is strongly against it.

To illustrate the impact of a prespecified significance level (α) on the probability of a Type II error (β), and hence on power ($1 - \beta$), let us consider situation A presented earlier in the chapter to illustrate key concepts involved in hypothesis testing. In situation A Karen, product manager for a line of apparel, is wondering whether to introduce the product line into a new market area. Karen does not want to introduce the product line unless she is reasonably sure that the mean household income in the market area is greater than \$12,000. Specifically, she wants to allow no more than a 5% chance of introducing the product line when the population mean income is actually less than or equal to \$12,000 (i.e., Karen wants to limit α, the probability of a Type I error, to .05). The null and

---

[1] Several interesting analogies highlighting the trade-off between Type I and Type II errors are discussed in Robert Hooke, *How to Tell the Liars from the Statisticians* (New York: Dekker, 1983), pp. 67–70. For a discussion of how the probability of a Type II error can be reduced when there are multiple tests of the same hypothesis in different settings, see Solomon Dutka, "Combining Tests of Significance in Test Marketing Research Experiments," *Journal of Marketing Research,* 21 (February 1984), pp. 118–119.

Exhibit A.17.1 *Sampling Distribution Curve When μ = \$12,000 and α = .05*



alternative hypotheses implied in situation A are as follows:

$$H_o: \mu \leq \$12{,}000;$$
$$H_a: \mu > \$12{,}000.$$

Exhibit A.17.1 contains the sampling distribution diagram centered at μ = \$12,000 and showing the critical sample mean value, $\bar{x}_c$, when $\alpha = .05$ (a sample size of 400 and standard deviation of \$8,000, the same values used in the chapter, have been assumed in computing the critical value). On the basis of the diagram in Exhibit A.17.1, Karen should reject H_o and introduce the product line if the sample mean exceeds \$12,658. Let us now examine the implication of this decision rule when H_o is actually false.

Suppose the true population mean (μ) is \$12,500 with the same standard deviation (and hence same shape of the sampling distribution curve) as before. Can a random sample of 400 households now yield a mean value of less than or equal to \$12,658? Yes, it can, as shown by Exhibit A.17.2(a).

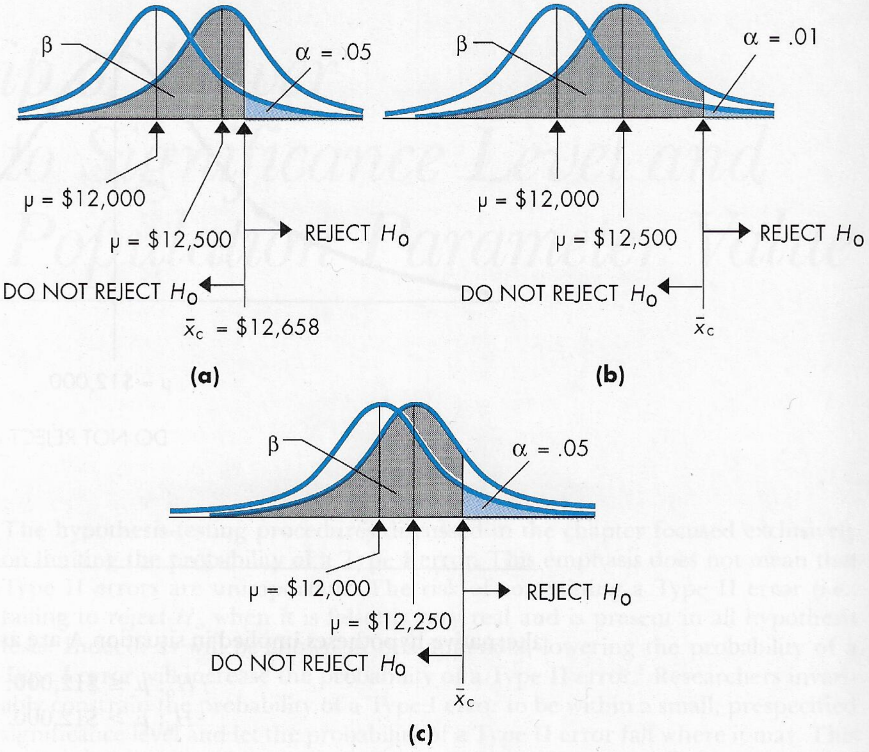
The shaded area under the sampling distribution curve centered on μ = \$12,500 in Exhibit A.17.2(a) is the probability of obtaining a sample mean less than or equal to \$12,658. Since, according to Karen's decision rule, H_o will *not* be rejected when $\bar{x} < \$12{,}658$, the shaded area is also the probability of a Type II error (β).

Although the actual values of β and power can be calculated in this case, we will not attempt to do so since the calculation procedure is quite cumbersome.[2]

---

[2] Most basic statistics texts discuss the procedure for calculating power. See, for example, Morris Hamburg, *Statistical Analysis for Decision Making* (New York: Harcourt Brace Jovanovich, 1977), pp. 271–275.

Exhibit A.17.2 *Impact of α and μ on β and the Power of the Test*



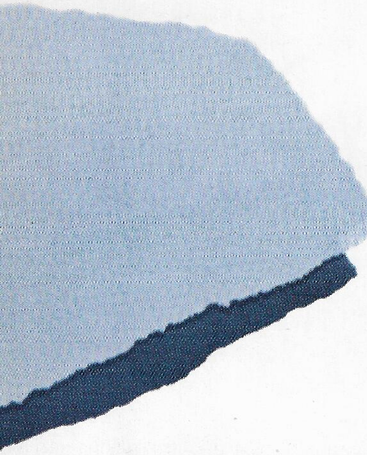
Rather, we will graphically examine the impact of changes in α and in the true population mean on β and power.

What will happen to the value of β in Exhibit A.17.2(a) if α is lowered from .05 to, say, .01? The impact of this change is shown in Exhibit A.17.2(b). Comparing Exhibits A.17.2(a) and A.17.2(b), we see that β increases and power decreases when α decreases. In other words, the probabilities of Type I and Type II errors are *inversely* related. (However, the sum of the α and β probabilities does *not* equal 1, as is often mistakenly believed.) This inverse relationship also makes intuitive sense: The smaller the value of α, the stronger is the evidence needed to reject H_o and, therefore, the greater are the chances of failing to reject H_o when it should be. The preceding point highlights the danger inherent in the blind choice of a small value for the significance level (e.g., .05): It implicitly assumes that the *opportunity cost* of making a no-go decision is negligible—or at least is not as severe as the cost of incorrectly making a go decision. Such an assumption may be questionable in many real-life settings.

Will β and power change if we assume that the true population mean is $12,250 instead of $12,500, as assumed earlier? A comparison of Exhibits A.17.2(a) and A.17.2(c) answers this question: With the changed assumption about the population mean, β increases and power decreases. This finding implies the following important generalization: When H_o is false, the values of β and power depend on how false H_o really is. The closer it is to being true (i.e., the closer the actual population mean is to the originally hypothesized mean), the higher β will be and the lower power will be. Therefore β and power depend not only on the prespecified significance level (i.e., α) but also on the actual (but unknown) value of the population parameter. This complexity of the concept of power is a key reason why hypothesis-testing procedures merely focus on limiting the probability of a Type I error, without explicitly incorporating the probability of a Type II error.

CHAPTER 18

Techniques for Examining Associations: Chi-Square, Correlation, & Regression



CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *State the purpose of the chi-square contingency test, perform such a test, and interpret the result.*
- *Determine the degree of association between variables in a contingency table.*
- *Compute the Spearman correlation coefficient between ordinal-scaled variables and determine whether or not it is statistically significant.*
- *Describe what the Pearson correlation coefficient represents and test its statistical significance.*
- *Distinguish between simple regression and correlation analyses.*
- *State the distinction between a dependent variable and an independent variable.*
- *Describe what the least squares approach does in determining a regression equation.*

- *Provide an interpretation of the terms Y-intercept and slope.*
- *Evaluate the trustworthiness of a regression equation.*
- *State and illustrate the possible uses of simple regression analysis.*
- *List the precautions to be taken in interpreting and using a simple regression equation.*
- *Define multicollinearity and indicate what it implies in terms of interpreting and using a multiple regression equation.*

The previous chapter covered several univariate techniques and hypothesis tests for investigating the statistical significance of *differences* between the frequency distributions/means/proportions pertaining to one variable. The focus of this chapter is on the investigation of *associations* between two or more variables. Table 18.1 lists the techniques discussed in this chapter, indicates when they are appropriate, and illustrates their potential applications. Certain other multivariate techniques are covered in the next chapter. The first technique we will look at is an extension of the two-way tabulation procedure discussed in Chapter 16.

CHI-SQUARE CONTINGENCY TEST

Chi-square Contingency Test *A nonparametric test for determining whether there is a statistically significant relationship between two categorical (nominal or ordinal) variables that have been cross-tabulated.*

The ***chi-square contingency test*** is a widely used technique for determining whether there is a statistically significant relationship between two categorical (nominal or ordinal) variables. (While the chi-square test only requires nominal data, it can also be used to analyze associations between two ordinal-scaled variables or one nominal- and one ordinal-scaled variable.) Recall that a mere visual inspection of a two-way tabulation of data can suggest whether or not the variables are associated with each other. The chi-square contingency test is a means of formally checking the relationship between such variables. It is a hypothesis test much like the chi-square goodness-of-fit test discussed in the previous chapter. To illustrate the chi-square contingency test, let us consider the following scenario.

Table 18.1 *Overview of Techniques for Examining Associations*

Technique	When the Technique Is Appropriate	Illustrative Research Question(s) the Technique Can Answer
Chi-square contingency test	When the association between two categorical (nominal or ordinal) variables is to be examined	Is there a significant relationship between households' income (measured, say, as a four-category ordinal variable) and whether or not they buy no-name-brand garbage bags?
Spearman correlation coefficient	When the degree of association between two sets of ranks (pertaining to two variables) is to be examined	Is there a significant relationship between motivation levels of salespeople and the quality of their performance? (Assume that the data on motivation and quality of performance are in the form of ranks, say, 1 through 20, for 20 salespeople who were evaluated subjectively by their supervisor on each variable.)
Pearson correlation coefficient	When the degree of association between two metric-scaled (interval or ratio) variables is to be examined	Is there a significant relationship between customers' age (measured in actual years) and their perceptions of our company's image (measured on a scale of 1 to 7)?
Simple regression analysis	When a mathematical function or equation linking two metric-scaled (interval or ratio) variables is to be constructed, under the assumption that values of one of the two variables is dependent on the values of the other	Is sales (measured in dollars) significantly affected by advertising expenditures (measured in dollars)? What proportion of the variation in sales is accounted for by variation in advertising expenditures? How sensitive are sales to changes in advertising expenditures?
Multiple regression analysis	Same as under simple regression analysis except that more than two variables are involved wherein one variable is assumed to be dependent on the others	Is sales significantly affected by advertising expenditures and price (where all three variables are measured in dollars)? What proportion of the variation in sales is accounted for by advertising and price? How sensitive are sales to changes in advertising and price?

Table 18.2 *Two-Way Tabulation of Working Status and Expenditures on Cosmetics*

Expenditures on Cosmetics	Working Status			Total
	Working Full-Time	Working Part-Time	Not Working	
Less than \$10	30 (21.4%)	20 (12.5%)	60 (30%)	110
\$10 to \$25	55 (39.3%)	60 (37.5%)	65 (32.5%)	180
Over \$25	55 (39.3%)	80 (50%)	75 (37.5%)	210
Total	140 (100%)	160 (100%)	200 (100%)	500

Example. The marketing manager of a cosmetics company is reviewing the results of a survey of adult women. The survey involved a random sample of 500 respondents and was conducted in a metropolitan area representative of the company's target market area. He is intrigued by one table, which is a cross-tabulation of data on the working status of respondents and the dollar amount per month they spend on cosmetics. Table 18.2 presents this cross-tabulation. Can the marketing manager infer that there is an association between working status and expenditure on cosmetics?

The percentage breakdowns in Table 18.2 do suggest an association between the two variables. And the association does appear to be somewhat intriguing, with those working part-time spending more on cosmetics than either of the other two categories of respondents. However, is this result trustworthy, or could the association have occurred in this sample purely by chance? A chi-square contingency test of the following hypotheses can answer this question:

H_o: There is no association between working status and expenditure on cosmetics (i.e., the two variables are independent of each other).

H_a: There is some association between working status and expenditure on cosmetics (i.e., the two variables are not independent of each other).

Conducting the Test

Contingency Table
A cross-tabulation set up for the purpose of conducting a chi-square contingency test.

Just as in the chi-square goodness-of-fit test, computing the test statistic in the chi-square contingency test involves comparing the actual, or observed, cell frequencies in the cross-tabulation (also called the ***contingency table*** in this context) with a corresponding set of expected cell frequencies. The expected cell frequencies are generated under the assumption that the null hypothesis is true. In other words, they are generated by asking the following question: "How would the total sample have partitioned itself into the various cells if the two variables were truly unrelated in the population?" The expected cell frequency (E_{ij}) for any cell defined by the *i*th row and *j*th column in the contingency table is given by

$$E_{ij} = \frac{n_i n_j}{n},$$

where n_i and n_j are the marginal frequencies—that is, the total number of sample units in category *i* of the row variable and category *j* of the column variable, respectively.

The rationale for this expression is based on probability theory and is quite intuitive. The probability that any given respondent within the sample will be in category *i* of the row variable is n_i/n. Similarly, the probability that this same respondent will be in category *j* of the column variable is n_j/n. Therefore, if the row and column variables are independent, the joint probability of the respondent being in both row category *i* and column category *j* is given by $(n_i/n)\,(n_j/n)$. For the total sample of *n* respondents the expected value (number of respondents) one should find in cell *ij* if H_o is true is given by $(n)(n_i/n)(n_j/n)$, which simplifies to $n_i n_j/n$.

Using this expression, we can compute an E_{ij}-value for each cell. For instance, the expected frequency for the first-row, first-column cell is given by

$$E_{11} = \frac{110 \times 140}{500} = 30.8.$$

Table 18.3 contains the expected cell frequencies corresponding to the various cells in Table 18.2. The row (as well as column) totals in

Table 18.3 *Observed and Expected Cell Frequencies for the Working Status/Cosmetics Expenditures Table*

Expenditures on Cosmetics	Working Status			
	Working Full-Time	Working Part-Time	Not Working	Total
Less than \$10	30 (30.8)	20 (35.2)	60 (44)	110
\$10 to \$25	55 (50.4)	60 (57.6)	65 (72)	180
Over \$25	55 (58.8)	80 (67.2)	75 (84)	210
Total	140	160	200	500

Note: In each cell *ij* the number without parentheses is the observed cell frequency (O_{ij}) and the number in parentheses is the expected cell frequency (E_{ij}).

Table 18.3 are identical for both the observed and expected frequencies. This feature is always true and is helpful in verifying whether the expected cell frequencies have been computed correctly.

The value of the *chi-square test statistic* in a contingency test is obtained by using the following formula:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

where r and c are the number of rows and columns, respectively, in the contingency table. Notice that this formula closely resembles the one we used for the chi-square goodness-of-fit test. The only difference between the two is that the expression here has a double summation sign and uses double subscripts because it relates to two categorical variables. Also, the number of degrees of freedom associated with this chi-square statistic are given by the product $(r - 1)(c - 1)$. In our example

$$\text{d.f.} = (3 - 1)(3 - 1) = 4.$$

Assuming a significance level of .05, the critical chi-square (χ^2_c) value from Appendix C for 4 degrees of freedom is 9.49. Furthermore, a chi-square contingency test of independence between two variables is always a one-tailed test. Therefore the decision rule is, "Reject H_o if $\chi^2 > 9.49$."

The computed value of the test statistic is

$$\begin{aligned}\chi^2 &= \frac{(30 - 30.8)^2}{30.8} + \frac{(20 - 35.2)^2}{35.2} + \frac{(60 - 44)^2}{44} \\ &\quad + \frac{(55 - 50.4)^2}{50.4} + \frac{(60 - 57.6)^2}{57.6} + \frac{(65 - 72)^2}{72} \\ &\quad + \frac{(55 - 58.8)^2}{58.8} + \frac{(80 - 67.2)^2}{67.2} + \frac{(75 - 84)^2}{84} \\ &= .021 + 6.564 + 5.818 + .420 + .100 + .681 \\ &\quad + .246 + 2.438 + .964 \\ &= 17.252.\end{aligned}$$

Since the computed chi-square value is greater than the critical value of 9.49, we can reject the null hypothesis. In other words, the apparent relationship between working status and expenditure on cosmetics revealed by the sample data is unlikely to have occurred because of chance. In fact, the *actual* significance level associated with a chi-square value of 17.252 is *less than* .005 (from Appendix C). Thus the chances of getting a chi-square value as high as 17.252 when there is no relationship between working status and expenditure on cosmetics are less than 5 in 1000.

That the chi-square contingency test only requires categorical data for ascertaining whether two variables are associated is a great advantage and has contributed to its wide use. However, for the test to be meaningful, there is a requirement for a minimum expected cell frequency (i.e., minimum E_{ij}). A commonly suggested rule of thumb is that no cell should have an expected frequency of less than one and no more than a fifth of the cells should have expected frequencies of less than five. The rationale here is similar to the rationale we discussed in the previous chapter for the chi-square goodness-of-fit test. Too many low expected frequencies will artificially inflate the computed chi-square value and may lead to the rejection of H_0 when such a conclusion is not warranted.

In practical terms, the minimum–expected frequency requirement implies that the test may not be meaningful when the observed *marginal* frequencies of one or both variables (i.e., the n_i- or n_j-values) are very small for certain categories. In some cases, however, one may be able to combine adjacent variable categories in such a way that the collapsed category has a large enough marginal frequency.

For example, suppose we want to conduct a chi-square contingency test to see whether there is an association between annual family income and the number of television sets owned by families; the data are gathered from a random sample of 250 families. Also, suppose family income is measured as a five-category variable and number of TVs owned is measured as a four-category variable; their observed marginal-frequency distributions are as shown in Table 18.4(a).

The last category of each variable has relatively few units, as shown in Table 18.4(a). The effect of these low observed frequencies on the expected cell frequencies is shown in Table 18.4(b). The last row–last column cell in this table has an expected frequency of only .2. Also, more than 20% of the cells in the table have expected frequencies of less than 5. The chi-square contingency test is therefore inappropriate for examining the association between the two variables in their present form.

Notice, however, that the categories of both variables are ordinal. Hence we can meaningfully combine the last two categories of each to form a single category with an adequate marginal frequency. The revised variable categories have large enough marginal totals to yield expected cell frequencies that satisfy the rule of thumb, as indicated in Table 18.4(c).

The ordinal nature of the variables in this example facilitated combining adjacent categories as desired. However, when a variable's categories do not follow a logical ordering—for example, those of race or ethnic background—combining categories may not always make sense. Furthermore, combining categories will result in a loss of degrees of freedom, which, in turn, will weaken the test. Another limitation of the chi-square contingency test is that it is inappropriate for examining the association between two measurements of the same variable taken at two

Table 18.4 *Association Between Family Income and Number of TVs Owned*

(a) Observed Marginal-Frequency Distributions

Family Income	Frequency	Number of TVs Owned	Frequency
1. Less than $10,000	100	1. One	80
2. $10,000 to less than $25,000	50	2. Two	90
3. $25,000 to less than $40,000	50	3. Three	70
4. $40,000 to less than $55,000	45	4. Four or more	10
5. $55,000 or more	5		
Total	250	Total	250

(b) Expected Cell Frequencies in a Two-Way Tabulation

Family Income Categories	Number of TVs Owned Categories 1	2	3	4	Total
1	32	36	28	4	100
2	16	18	14	2	50
3	16	18	14	2	50
4	14.4	16.2	12.6	1.8	45
5	1.6	1.8	1.4	0.2	5
Total	80	90	70	10	250

(c) Expected Cell Frequencies with Collapsed Variable Categories

Family Income Categories	Number of TVs Owned Categories 1	2	3+	Total
1	32	36	32	100
2	16	18	16	50
3	16	18	16	50
4+	16	18	16	50
Total	80	90	80	250

different times (e.g., the brand preferences of a consumer panel before and after a major promotional campaign for one of the brands).[1] Finally, the chi-square contingency test will only indicate whether or not there is a statistically significant relationship between two variables, without providing a meaningful quantitative measure of the strength of the relationship. This limitation can be overcome to some extent by computing certain additional statistics, as described in the next section.

Measuring the Degree of Association in Contingency Tables

Contingency Coefficient
An index that measures the degree of association between two variables in a contingency table when the null hypothesis of independence between the variables is rejected.

When the null hypothesis of independence between variables is rejected in a chi-square contingency test, some feeling for the degree of association between the two variables can be obtained by computing a ***contingency coefficient*** (C) as follows:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}},$$

where χ^2 is the computed chi-square value and n is the total sample size. In the example pertaining to Table 18.3,

$$C = \sqrt{\frac{17.252}{500 + 17.252}} = .183.$$

The lowest possible value of C is zero, which will occur when $\chi^2 = 0$ (i.e., when there is absolutely no relationship between the two variables). Unfortunately, the upper limit on C is not fixed; it depends on the number of rows and columns in the contingency table and is not necessarily unity. Interpreting the computed value of C is therefore not easy.[2]

Cramer's V-Statistic
An index (varying between 0 and 1) that measures the degree of association between two variables in a contingency table when the null hypothesis of independence between the variables is rejected. This index is a somewhat more useful measure than the contingency coefficient.

When the number of rows and the number of columns in a contingency table are equal, however, the maximum value of C is given by $\sqrt{(r-1)/r}$, where r is the number of rows (or columns). Since the contingency table in our example has three rows and three columns, the maximum possible value of C is $\sqrt{\frac{2}{3}}$, or .816. This value suggests that the association between working status and expenditure on cosmetics is quite weak; the C-value is only .183 when its range is from 0 (no relationship) to .816 (perfect relationship).

Given the difficulty in interpreting the contingency coefficient, especially when the number of rows and columns in the table are different, a somewhat more useful measure, called ***Cramer's V-statistic***, has been proposed. The expression for this statistic is

---

[1] One nonparametric test that can be used to measure the significance of a change between a before and an after measure of the same variable is the McNemar test; for details, see Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill, 1956), pp. 63–67.

[2] For further discussion of the difficulty in interpreting the contingency coefficient, see Siegel, *Nonparametric Statistics*, pp. 200–201.

$$V = \sqrt{\frac{\chi^2}{n(f - 1)}},$$

where χ^2 and n are defined as before and f is the smaller of the number of rows and columns in the contingency table, that is, $f = \min(r, c)$. The quantity $n(f - 1)$ is the maximum value that the computed χ^2 can possibly attain for any $r \times c$ contingency table. Therefore $V = 0$ when there is no relationship and $V = 1$ when there is a perfect relationship. Hence, at least from an intuitive standpoint, V is a better measure of degree of association than the contingency coefficient C.[3] In our example

$$V = \sqrt{\frac{17.252}{500(3 - 1)}} = .131.$$

This value of V is consistent with our earlier inference that the degree of association between the two variables is quite weak.

A word of caution about the contingency coefficient and Cramer's V-statistic is in order. Although they seem like precise measures of association between variables, there is a certain amount of arbitrariness built into them. That is, the number of categories created for each variable and how those categories are defined can greatly influence the configuration of the contingency table—and hence the values of these statistics—for the *same* set of raw data.

To illustrate, consider Table 18.5(a). Categorical data on two variables, age and brand of soft drink most preferred, obtained from a sample of 90 respondents are cross-tabulated in Table 18.5(a). This table implies a perfect association between the two variables. Therefore the contingency coefficient C will attain its maximum value of .816 (since the table has three rows and three columns), and Cramer's V-statistic will also attain its maximum value of 1. Let us assume for a moment that 15 of the 30 respondents in the 20–35 age category are 35 years old, and that 15 of the 30 respondents in the 36–50 age category are 50 years old. Also, suppose the following slightly modified age categories are used in constructing the contingency table:

20–34 (instead of 20–35)
35–49 (instead of 36–50)
50 and over (instead of over 50)

With the redefined age categories, the cross-tabulation of the same data will look like the contingency table shown in Table 18.5(b). This table clearly does not suggest as perfect a relationship between the two

---

[3] For additional details, see, for example, Jean Dickinson Gibbons, *Nonparametric Methods for Quantitative Analysis* (New York: Holt, Rinehart and Winston, 1976), pp. 330–339.

Table 18.5 *Association Between Age and Brand of Soft Drink Most Preferred*

(a) Original Contingency Table

	Most Preferred Brand			
Age	**A**	**B**	**C**	**Total**
20–35	30	0	0	30
36–50	0	30	0	30
Over 50	0	0	30	30
Total	30	30	30	90

(b) Contingency Table with Revised Age Categories

	Most Preferred Brand			
Age	**A**	**B**	**C**	**Total**
20–34	15	0	0	15
35–49	15	15	0	30
50 and over	0	15	30	45
Total	30	30	30	90

variables as does Table 18.5(a). Therefore, the contingency coefficient and Cramer's *V*-statistic corresponding to Table 18.5(b) will be lower than their maximum possible values (you can verify this result).

In short, numbers representing the degree of association on the basis of a contingency table can be very sensitive to how the table is constructed. They are hence not as precise as measures of association for metric-scaled variables (called correlation coefficients, which we will discuss a little later). After all, as we saw earlier, analysis of nonmetric data cannot yield as rich and refined results as analysis of metric data.

SPEARMAN CORRELATION COEFFICIENT

Spearman Correlation Coefficient *An index (varying between −1 and +1) that measures the nature and the degree of association between two ordinal variables.*

The chi-square contingency test and the corresponding *C*- and *V*-statistics can be used for examining the association between variables whose categories are ordinal. However, a more powerful measure of association between two ordinal variables, called the ***Spearman correlation coefficient*** (r_s), can be computed when the data are refined enough to rank the sample units 1 through *n* on each variable. Consider the following example.

Example. Over the past several years an industrial marketing firm has been hiring all its salespeople from among the graduates of 10

Table 18.6 *Association Between School Prestige and Performance of Graduates*

Business School (i)	Ranking of School's Prestige (SP_i)	Ranking of Performance of School's Graduates (GP_i)	Difference Between Ranks ($d_i = SP_i - GP_i$)	Squared Difference (d_i^2)
1	10	8	2	4
2	7	3	4	16
3	9	7	2	4
4	1	2	−1	1
5	6	9	−3	9
6	2	4	−2	4
7	3	5	−2	4
8	8	10	−2	4
9	5	6	−1	1
10	4	1	3	9
				$\sum_{i=1}^{10} d_i^2 = 56$

business schools in the vicinity of its headquarters. To ascertain whether there is any association between the relative prestige of the 10 schools and the performance of the graduates hired from each, the firm's sales managers developed a subjective ranking of the perceived prestige levels of the 10 schools and the performance levels of the groups of graduates recruited from these schools. These rankings are shown in the second and third columns of Table 18.6. What is the degree of association between the prestige levels of the schools and the sales performance levels of their graduates hired by this company?

A Spearman correlation coefficient is appropriate for answering the question posed in this situation. It is a widely used measure of association between two sets of ranks. The formula for computing it is

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)},$$

where d_i is the difference between the ith sample unit's ranks on the two variables and n is the total sample size. Notice that $r_s = 1$ when the sum of the d_i^2 values is zero, that is, when the two sets of ranks are identical.[4] The range of possible values for r_s is $+1$ (perfect direct association) to -1 (perfect inverse association), with a value of 0 signifying no association.

As shown in the last column of Table 18.6, the value of $\Sigma\, d_i^2$ in our example is 56. Therefore

$$r_s = 1 - \frac{(6)(56)}{10(100 - 1)} = 1 - .339 = .661.$$

The r_s-value of .661 suggests at least a moderate association between the two sets of ranks. Is this association statistically significant? Answering this question involves assuming that the 10 schools are a random sample from a population of business schools and testing these hypotheses:

$$H_o\colon \rho_s = 0;$$
$$H_a\colon \rho_s \neq 0;$$

where ρ_s is the population correlation coefficient between the two sets of ranks. When $n \geq 10$, the following *test statistic* will have a *t-distribution* with $n - 2$ degrees of freedom:

$$t = r_s \sqrt{\frac{n - 2}{1 - r_s^2}}.$$

This hypothesis test is a two-tailed test. Assuming a significance level of .05, the critical values of t for 8 degrees of freedom (d.f. $= n - 2 = 10 - 2 = 8$) are $+2.31$ and -2.31. Hence the decision rule is, "Reject H_o if $t > 2.31$ *or* if $t < -2.31$." In the present example

$$t = .661 \sqrt{\frac{10 - 2}{1 - (.661)^2}} = .661 \sqrt{\frac{8}{1 - .437}}$$
$$= (.661)(3.770) = 2.49.$$

Since $t > 2.31$, we reject H_o and conclude that there is a true association between the prestige of business schools and the job performance of its graduates. In other words, the sample correlation of .661 is unlikely to have occurred because of chance.

Note that the basis for determining r_s is the computation of differences between pairs of ranks, an operation that is apparently inconsistent with the notion that it is inappropriate to place any meaning on the magnitude of the difference between numbers forming only an ordinal scale. Indeed, in computing r_s one is implicitly assuming that the differences between ranks on the two variables can be meaningfully compared. If such an assumption is unreasonable, the Spearman correlation coefficient may be misleading.

---

[4] For a discussion of the derivation and interpretation of the formula for r_s, see Siegel, *Nonparametric Statistics*, pp. 202–213. This text also discusses the treatment of tied ranks.

PEARSON CORRELATION COEFFICIENT

Pearson Correlation Coefficient (Product-Moment Correlation) *An index (varying between −1 and +1) that measures the nature and the degree of association between two metric-scaled (interval or ratio) variables.*

The ***Pearson correlation coefficient*** (also known as the ***Pearson product-moment correlation***) is appropriate for measuring the degree of association between variables that are interval- or ratio-scaled. It is a more refined measure than the ones we have discussed so far. The Pearson correlation coefficient is insightful in its own right and also plays a key role in advanced multivariate analysis procedures.

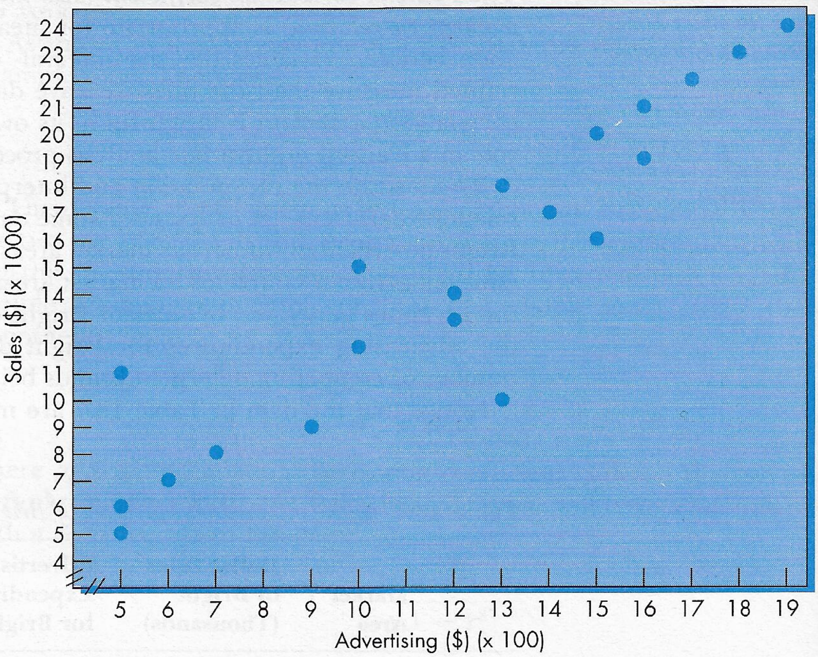
To illustrate the computation and interpretation of the Pearson correlation coefficient, let us consider some data gathered by a firm selling Bright detergent in numerous market areas nationwide. Table 18.7 contains data from a sample of 20 market areas for the following variables: the revenues generated by sales of Bright during a one-month period, the advertising expenditures for Bright during that period, and the number of competing detergent brands being sold in each market area.

Notice that the data in Table 18.7 are metric data; indeed, all three

Table 18.7 *Data Gathered by Marketer of Bright Detergent*

Market Area	Dollar Sales of Bright (Thousands)	Advertising Expenditures for Bright ($)	Number of Competing Detergents
1	5	500	15
2	10	1300	8
3	6	500	14
4	20	1500	5
5	15	1000	9
6	9	900	10
7	11	500	12
8	18	1300	4
9	22	1700	6
10	7	600	13
11	24	1900	2
12	14	1200	8
13	16	1500	6
14	17	1400	7
15	23	1800	1
16	8	700	11
17	12	1000	10
18	13	1200	7
19	21	1600	7
20	19	1600	3

Exhibit 18.1 *Scatter Diagram of Sales and Advertising Data*



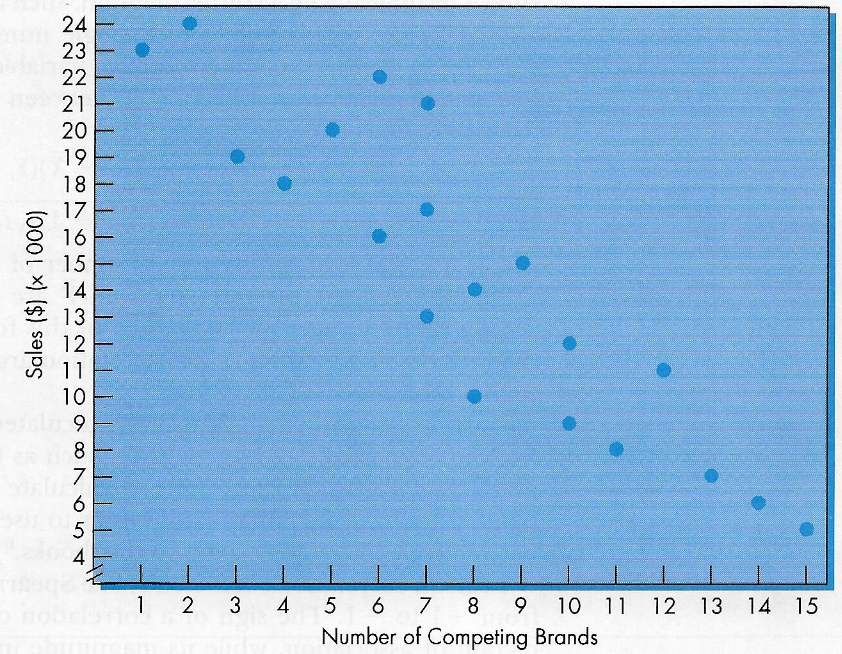
variables are ratio-scaled. Therefore as a first step in understanding the association between any two of the variables, we can plot the corresponding data on a two-dimensional graph. Such a plot is called a ***scatter diagram***, and it can indicate how closely and in what fashion the variables are associated. Exhibit 18.1 shows a scatter diagram of the sales and advertising data; Exhibit 18.2 shows a plot of sales versus number of competing brands.

Scatter Diagram
A two-dimensional graph in which data on two metric-scaled (interval or ratio) variables are plotted.

The scatter diagram in Exhibit 18.1 clearly shows a direct relationship between dollar sales and advertising expenditures: In general, the higher the advertising expenditures, the higher the sales are. Also, the swarm of points in the diagram seems to follow a *linear trend*. We will have more to say about this linear trend later when we discuss regression analysis. However, we note here that the existence of a linear scatter is a prerequisite for the meaningfulness of the Pearson correlation coefficient, because it is only designed to capture the extent of linear association between variables.

Two other key assumptions that must be met, especially for making statistical inferences based on the Pearson correlation coefficient, are the following:

Exhibit 18.2 *Scatter Diagram of Sales and Number of Competing Brands Data*



1. The two variables have a bivariate normal distribution. That is, the population is such that all units with a given value of one variable have values on the second variable that are normally distributed.
2. The variance of the normal distribution of one variable at any given value of the other variable remains the same across various values of the latter variable.

These assumptions are hard to verify by using sample data and are often taken for granted.[5]

The scatter diagram in Exhibit 18.2 also displays an underlying linear association between sales and number of competing brands. Notice, however, that this trend is an inverse one; that is, in general, the greater the number of competing brands, the lower the sales of Bright are. Exhibit

---

[5] Further discussion of these assumptions can be found in many statistics textbooks; see, for example, David R. Anderson, Dennis J. Sweeney, and Thomas A. Williams, *Introduction to Statistics: An Applications Approach* (St. Paul, Minn.: West, 1981), pp. 343–345.

18.2 thus suggests a fairly strong, but negative, association between sales and number of competing brands.

The direction and the degree of association between two variables shown graphically by a scatter diagram, such as the ones in Exhibits 18.1 and 18.2, can be quantified in a single number, which is the Pearson correlation coefficient. When the two variables are labeled X and Y, the Pearson correlation coefficient (r_{XY}) between them is given by

$$r_{XY} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{(n - 1)s_X s_Y},$$

where n is the sample size (total number of data points), X_i and Y_i are values for any sample unit i, $\overline{X}$ and $\overline{Y}$ are means, and s_X and s_Y are standard deviations. The derivation of this formula and a discussion of the rationale underlying its interpretation are provided in the appendix to this chapter.

Correlation coefficients are easily calculated by using readily available computer packages once a data set (such as the one in Table 18.7) has been created. But if one wants to calculate correlation coefficients by hand, a shortcut formula that is easier to use than the expression given here is available in basic statistics textbooks.[6]

Pearson correlation coefficients, like Spearman coefficients, can range from $+1$ to -1. The sign of a correlation coefficient indicates the direction of association, while its magnitude indicates the strength of association. In our example the correlation between sales and advertising turns out to be .927. This value reaffirms the strong positive association suggested by the scatter diagram in Exhibit 18.1. Consistent with our interpretation of the scatter diagram in Exhibit 18.2, the correlation between sales and number of competing brands is $-.910$ (a strong inverse association).

The sample correlation coefficient r can itself serve as the test statistic in conducting a two-tailed test of the following hypotheses:

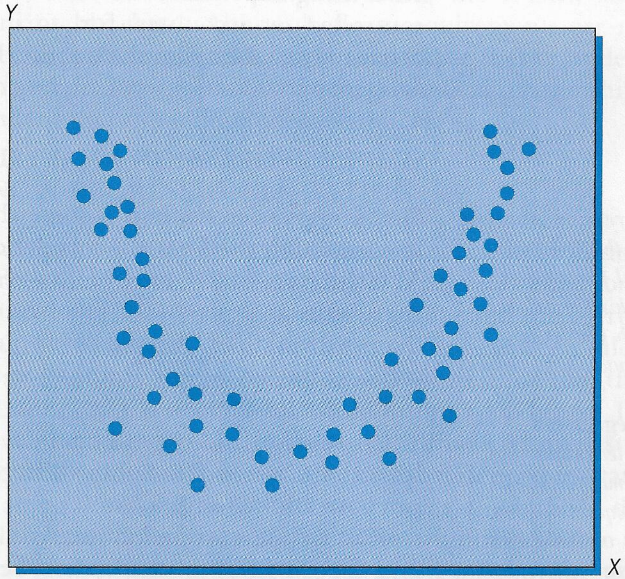
$$H_o: \rho = 0;$$
$$H_a: \rho \neq 0;$$

where ρ is the population correlation coefficient. The degrees of freedom associated with r are $n - 1$, and its critical values are tabled in Appendix E. In our example, for $\alpha = .05$ and 19 degrees of freedom (d.f. $= n - 1 = 19$), the critical values are $r_c = +.433$ and $r_c = -.433$. The decision rule is, "Reject H_o if $r > .433$ *or* if $r < -.433$." Applying this decision rule to the .927 correlation between sales and advertising and to the $-.910$ correlation between sales and number of competing brands, we should reject H_o in both cases.

---

[6] See, for example, Anderson, Sweeney, and Williams, *Introduction to Statistics*, p. 357.

Exhibit 18.3 *Scatter Diagram Showing a Nonlinear Association Between Variables*



While Pearson correlation coefficients are extremely helpful in uncovering bivariate associations, there are two important caveats concerning their use. First, a low sample correlation (or a failure to reject H_O) does not necessarily mean there is *no* association; it only implies an absence of *linear* association. Exhibit 18.3 illustrates this point. The scatter diagram in this figure clearly shows a strong U-shaped relationship between X and Y. Yet the Pearson correlation coefficient in this case will be close to zero, since there is no discernible linear trend when the data points are considered together. The moral: Even if the Pearson correlation coefficient is very small or not statistically significant, one should explore the possible presence of a *nonlinear* association, especially when intuitive or theoretical considerations suggest that the variables may be related. Perhaps the simplest way of exploring a nonlinear association is to plot and visually examine a scatter diagram.

Second, as has been pointed out earlier in the book, existence of a correlation by itself is not sufficient to infer causation between variables. For instance, the correlation of .927 between sales and advertising does not necessarily imply that increased advertising expenditures will lead to increased sales, although it is tempting to draw such a conclusion. The only thing this correlation says is that the data on the two variables follow a very similar pattern in the 20 markets studies. But this pattern may be a result of a causal linkage going from sales to advertising, rather than the other way around. For example, anticipation of a certain level of sales in each market may have dictated a correspondingly low or high advertising expenditure level. Thus unless all other relevant factors are controlled or accounted for, to infer a certain direction of causality between variables simply on the basis of a correlation coefficient is dangerous.

SIMPLE REGRESSION ANALYSIS

Simple Regression Analysis *A technique that generates a mathematical function or equation linking a dependent variable with* **one** *independent variable.*

Multiple Regression Analysis *A technique that generates a mathematical function or equation linking a dependent variable with* **two or more** *independent variables.*

Simple regression analysis is somewhat similar to the procedure involved in computing the Pearson correlation coefficient between two variables. A technique called ***multiple regression analysis*** is not unlike simple regression analysis, but it involves examining associations among more than two variables. We will cover multiple regression analysis in the next section after laying the foundation for the general regression technique in this section.

Correlation and regression analyses have a lot in common. Nevertheless, there are a few subtle differences between the two in terms of purpose and assumptions.

First, while correlation analysis focuses on summarizing the degree and the direction of association between variables as a single number, the purpose of regression analysis is to generate a mathematical function or equation linking those variables. The resulting equation is called a *regression equation.*

Second, when we are examining the association between two variables, the notion of a presumed causal variable is much more germane to regression analysis than to correlation analysis. (Recall that we introduced the concept of a presumed causal variable in Chapter 16 while discussing two-way tabulations.) Specifically, in simple regression analysis we must designate one of the two variables as the ***independent variable***, which is also known as an ***explanatory*** or ***predictor variable***. The variable so designated is often presumed to be a cause of the other, and it is conventionally plotted along the X-axis in a scatter diagram. The second variable—the one presumed to be influenced by the independent variable—is appropriately called the ***dependent*** (or sometimes ***criterion***) ***variable*** and it is traditionally plotted along the Y-axis.

Independent (Explanatory or Predictor) Variable *A variable in multivariate analysis that is presumed to influence a dependent variable. It is conventionally plotted along the X-axis in a scatter diagram.*

Dependent (Criterion) Variable *A variable in multivariate analysis that is presumed to be influenced by one or more independent variables. It is conventionally plotted along the Y-axis in a scatter diagram.*

Third, while the assumptions about statistical properties of the input data necessary for correlation analysis are, in general, also required for regression analysis, there is one major difference: In regression analysis only the dependent variable is random; the independent variable is implicitly treated as a fixed variable. Stated differently, in generating the input data for regression analysis, one is supposed to set the independent variable (being presumed as a causal factor) at a variety of specific or fixed levels, at each of which one observes the nature of the distribution of values for the dependent variable. Furthermore, the dispersion (or variance) of dependent-variable values at each level for the independent variable should be approximately the same. That is, if one were to measure the width of a scatter diagram (i.e., swarm of points) in the vertical direction at various locations along the X-axis, such widths should be roughly equal. Although these requirements are not always strictly met in practice, they are important to the theory underlying the derivation and the interpretation of a mathematical relationship between the dependent and the independent variables.[7]

DERIVING A REGRESSION EQUATION

A scatter diagram is useful for developing a conceptual understanding of the process involved in constructing a regression equation. We will once again use the data on sales and advertising and the corresponding scatter diagram in Exhibit 18.1 for this purpose. Between sales and advertising, suppose management has reason to presume that the latter is the likely causal variable; that is, advertising expenditures for Bright detergent in the 20 market areas were not influenced by past or anticipated sales of Bright. Advertising is hence the independent variable and is plotted along the X-axis in Exhibit 18.1. Sales, the implied dependent variable, is plotted along the Y-axis. The resulting scatter diagram's linear trend can be summarized by constructing an appropriate upward-sloping straight line through it. Any straight line constructed on a two-dimensional graph can be mathematically represented by the following type of equation linking the X- and Y-variables:

$$Y = a + bX,$$

where a and b are constants. An equation of this form is what regression analysis of data is intended to generate.

How does one construct an appropriate straight line through the scatter diagram? One obvious approach is to visually scan the scatter diagram and subjectively construct a line that appears to fit it well. The problem with this approach, however, is that it is unlikely to produce one unique line; several different lines may all seem to fit the scatter diagram equally well, as Exhibit 18.4 demonstrates. Clearly, then, a more objective procedure is needed to identify *the* best-fitting line. One such widely used procedure, called the ***least squares approach***, is discussed in the appendix to this chapter.

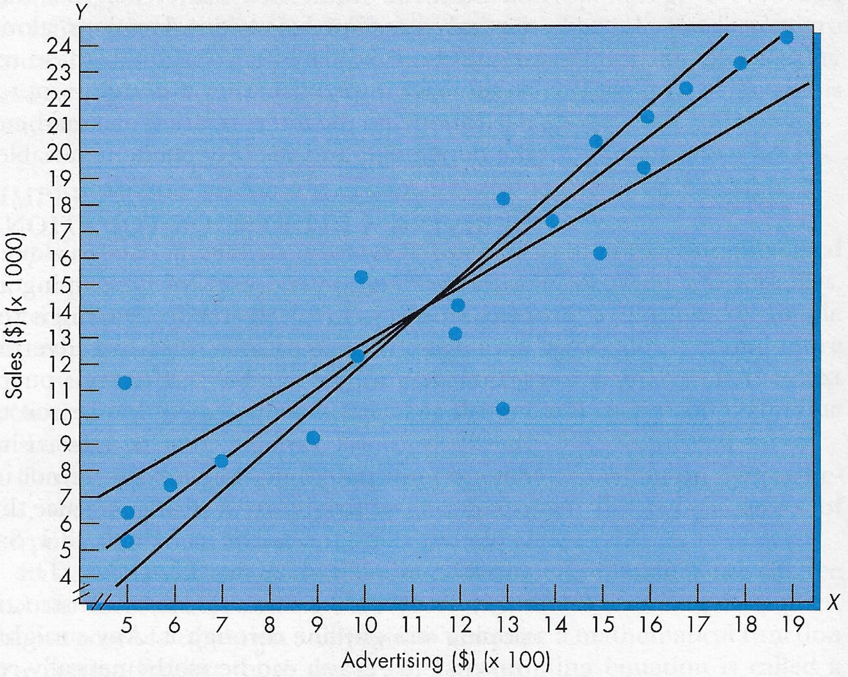
Least Squares Approach
An analysis procedure widely used in deriving a regression equation that is the best-fitting equation for a given set of data on dependent and independent variables.

Statistical packages available on most computers are capable of performing least-squares regression analysis. They can provide the values of a and b as well as a variety of statistics for evaluating the regression

---

[7] For further discussion of the rationale for this requirement and the extent to which one can deviate from it without reducing the meaningfulness of the resulting regression equation, consult an advanced multivariate statistics book, such as J. Johnston, *Econometric Methods* (New York: McGraw-Hill, 1984).

Exhibit 18.4 *Several Subjectively Constructed Regression Lines*



equation. Exhibit 18.5 shows the computer output (produced by an SPSS program) for the regression analysis of the sales and advertising data.

We have not yet discussed all the regression analysis concepts necessary for fully understanding this output. Hence we will be making reference to this exhibit as and when we cover each concept. Here, let us simply get the values of *a* and *b* from the exhibit. These values are shown under the column labeled B in the lower half of the output. Specifically, the number corresponding to ADVTG (the name given to the advertising variable in this program) in this column is the value of *b*. The number immediately below that number is the value of *a*. Rounding these values to three decimal places, we have

$$a = .163,$$
$$b = 1.210.$$

The regression equation is

$$\hat{Y}_i = .163 + 1.210X_i,$$

where $\hat{Y}_i$ is the value of sales (in thousands of dollars) *predicted* by the regression equation for an advertising expenditure level of X_i (in hundreds of dollars).

Exhibit 18.5 *SPSS Computer Output for Simple Regression Analysis of Sales and Advertising Data*

```
                                   MULTIPLE REGRESSION

DEPENDENT VARIABLE. .      SALES                                              VARIABLE LIST 1
VARIABLE(S) ENTERED ON STEP NUMBER 1. .      ADVTG                          REGRESSION LIST 1

MULTIPLE R              0.92716   ANALYSIS OF VARIANCE   DF   SUM OF SQUARES   MEAN SQUARE          F
R SQUARE                0.85962   REGRESSION              1.       571.64576     571.64576  110.22127
ADJUSTED R SQUARE       0.85182   RESIDUAL               18.        93.35424       5.18635
STANDARD ERROR          2.27736

            VARIABLES IN THE EQUATION                        VARIABLES NOT IN THE EQUATION

VARIABLE            B          BETA       STD ERROR B       F      VARIABLE  BETA IN  PARTIAL  TOLERANCE  F
ADVTG          1.209832      0.92716        0.11524    110.221
(CONSTANT)     0.1634874
MAXIMUM STEP REACHED
STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.
```

Y-Intercept *The constant a in a regression equation; it represents the predicted value of the dependent variable when all independent variables are set to zero.*

Slope *The regression coefficient (b-value) corresponding to an independent variable in a regression equation; it represents the change in the predicted value of the dependent variable per one-unit change in the independent variable, assuming all other variables likely to influence the dependent variable remain the same.*

When $X_i = 0$, $\hat{Y}_i = .163$, which is the constant a. The a in a regression equation is known as the ***Y-intercept*** and represents the predicted value of the dependent variable corresponding to a value of zero for the independent variable.

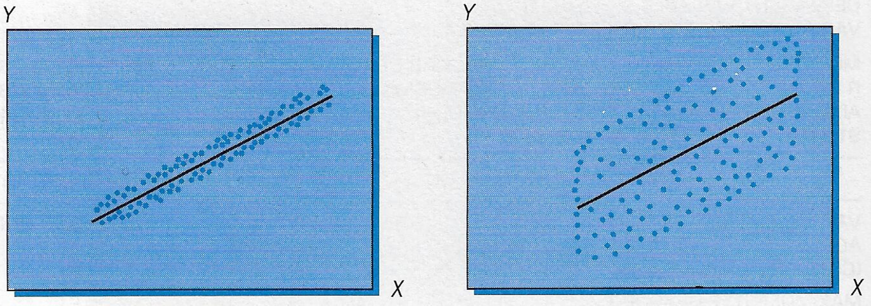
The constant b, representing the coefficient of X_i, is known as the ***slope*** of the regression equation. The slope has a very important interpretation: *It represents the change in the predicted value of the dependent variable per one-unit change in the independent variable*, assuming that all other variables likely to influence the dependent variable remain the same. In our example, since sales are measured in thousands of dollars and advertising expenditures are measured in hundreds of dollars, the value of 1.210 for b implies the following: If advertising expenditures are increased (or decreased) by \$100 (which constitutes one unit of advertising), sales can be predicted to increase (or decrease) by \$1210 (which constitutes 1.210 units of sales), assuming no other relevant factor changes.

EVALUATING THE REGRESSION EQUATION

As mentioned earlier, the least squares procedure will yield the best possible straight line and corresponding regression equation for a given data set. However, when the data observed in a scatter diagram are widely dispersed, as opposed to falling within a relatively narrow linear band that follows an upward- or downward-sloping trend, the usefulness of the resulting regression equation is questionable. Stated differently, the least squares procedure will always yield a regression equation; but how trustworthy the equation is depends on how compact the scatter diagram is and how closely it resembles a linear trend.

To illustrate, consider the two scatter diagrams shown in Exhibit 18.6.

Exhibit 18.6 *Two Identical Regression Lines Based on Different Scatter Diagrams*



(a) (b)

Suppose a least-squares analysis of either diagram results in the same regression equation (i.e., same *a*- and *b*-values). But, intuitively, the regression equation stemming from the scatter diagram of Exhibit 18.6(a) will be more trustworthy and valuable than one stemming from the diagram of Exhibit 18.6(b). Clearly, then, we need objective criteria for evaluating the *goodness* of the regression equation. One such criterion is called the ***coefficient of determination***, denoted by R^2. The coefficient of determination is a global measure of how much better predictions of the dependent variable made with the aid of the regression equation are than those made without.

Coefficient of Determination (R^2) *A global measure (varying between 0 and 1) of how good a regression equation is. It represents the proportion of variation in the dependent variable accounted for, or explained by, the independent variable(s) in the equation,*

For a given value of X, what is the best prediction of Y *without* using the regression equation? The best prediction should be $\overline{Y}$, the mean value of the dependent variable. While $\overline{Y}$ will greatly underpredict some actual values and greatly overpredict others, on the average (over a number of such predictions), the overpredictions and underpredictions will cancel out.

For any given value of X_i the deviation between the actual and the mean values of Y—that is, $(Y_i - \overline{Y})$—is called the *total deviation in* Y. The sum of the total squared deviations—that is, $(Y_i - \overline{Y})^2$ added across all data points in the scatter diagram—is a measure of the variation in Y about its mean value. This sum is conventionally called the ***total sum of squares***, denoted by SS_T. Part of SS_T can be accounted for, or explained, when an appropriate regression equation is available. The variance thus explained is called the ***regression sum of squares***, SS_R. The variance left unexplained by the regression equation is called the ***error***, or ***residual, sum of squares***, SS_E. It can be shown that $SS_T = SS_R + SS_E$. The procedure for doing so is called partitioning the total sum of squares, which is illustrated briefly in the chapter appendix.

Total Sum of Squares *Quantity obtained by adding across all data points the squared deviation of each actual value of the dependent variable from its mean value.*

Regression Sum of Squares *The portion of the total sum of squares that is explained by a regression equation.*

Error (Residual) Sum of Squares *The portion of the total sum of squares that is left unexplained by a regression equation.*

The least squares procedure fits a straight line through the scatter diagram such that SS_E has the smallest possible value. Since SS_T is fixed by the input data on Y, the SS_R corresponding to the best-fit line will have the highest possible value. Intuitively, the higher the value of SS_R for a given value of SS_T, the better the regression equation should be. Therefore a good way to see whether a regression equation is trustworthy is to examine the ratio between SS_R and SS_T, which is exactly what the coefficient of determination R^2 accomplishes:

$$R^2 = \frac{\text{variance explained by the regression equation}}{\text{total variance}} = \frac{SS_R}{SS_T}$$

The right-hand half of the computer output in Exhibit 18.5 contains the values of SS_R and SS_E under the heading SUM OF SQUARES. Rounding the values to three decimal places, we obtain

$$SS_R = 571.646$$

$$SS_E = 93.354$$

$$SS_T = SS_R + SS_E = 665$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{571.646}{665} = .860$$

Thus from the regression equation in our example, advertising expenditures can be said to explain about 86% of the total variance in sales.

The range of possible values for R^2 is from 0 to 1. The R^2 for a regression equation corresponding to a scatter diagram that displays no apparent upward- (or downward-) sloping linear trend will be close to zero. When the scatter diagram forms a perfectly horizontal swarm of points, the regression line will coincide with a horizontal line constructed at $Y = \bar{Y}$. Predictions based on such a regression line will, of course, be no better than those given by $\bar{Y}$, and hence R^2 will be zero. What will R^2 be when all the data points in a scatter diagram fall exactly on a downward- or upward-sloping straight line? In such a case SS_E will be zero, SS_R will be the same as SS_T, and R^2 will be 1.

The value of R^2 we computed as the ratio SS_R/SS_T is also shown in Exhibit 18.5, next to the label R SQUARE. The ADJUSTED R SQUARE value immediately below it is a corrected value that takes into account the number of data points available for constructing the regression line vis-à-vis the number of independent variables. The adjusted R^2 will invariably be lower than the original R^2. Discussion of why and how this correction is made can be found in any advanced book on regression analysis.[8] Hereafter, we will examine only the unadjusted R^2-values because they have a more straightforward interpretation.

---

[8] See, for example, Ronald J. Wonnacott and Thomas H. Wonnacott, *Econometrics* (New York: Wiley, 1970), p. 311.

Multiple Correlation Coefficient *A measure of the overall degree of association between dependent and independent variables in a regression equation; it is simply the square root of the coefficient of determination (R^2).*

The item labeled MULTIPLE R in the computer output is simply the square root of R^2. It is generally known as the ***multiple correlation coefficient*** and represents the overall degree of association between the dependent and independent variables. Since there are only two variables in our example, the multiple correlation coefficient of .927 is equivalent to the Pearson correlation coefficient between sales and advertising. In any bivariate regression analysis the square of the Pearson correlation coefficient between the dependent and independent variables will be the same as the value of the coefficient of determination.

Testing the Significance of R^2

The R^2-value is an important measure of how trustworthy a regression equation is: The higher the R^2-value, the more trustworthy the equation is. Nevertheless, the regression equation and its R^2-value are based only on sample data. Thus depending on the nature of the data, one may sometimes obtain a high R^2-value when there is really no association between the dependent and independent variables in the *population*. For instance, suppose data on sales and advertising expenditures are obtained from a sample of just two markets chosen at random from a population of all such markets. A scatter diagram of these data will consist of just two points. We can fit a straight line through these points no matter where they are on the two-dimensional graph. Thus R^2 will be 1 since the fit will be perfect. But we cannot trust this regression line to be an accurate representation of any underlying association between sales and advertising. Thus a high R^2-value by itself is necessary *but not sufficient* to conclude that a regression equation is trustworthy.

A suitable test is needed to verify whether a nonzero R^2-value stemming from a sample regression line is truly reflective of an underlying linear association in the population. Whether a sample R^2-value is significantly greater than zero can be checked by computing the following *F-statistic*:

$$F = \frac{\mathrm{SS_R}/k}{\mathrm{SS_E}/(n - k - 1)},$$

where k is the number of independent variables included in the equation and n is the sample size (i.e., the number of data points).

To determine the critical value of F, we need to specify two types of degrees of freedom associated with the F-statistic: the numerator degrees of freedom, given by k, and the denominator degrees of freedom, given by $n - k - 1$. Critical values of F are tabled in Appendix F for the customary significance level (α) of .05. In our example $k = 1$ and $n - k - 1 = 18$. For these numerator and denominator degrees of freedom, the critical value of F from Appendix F is 4.41. The decision rule implied in this F-test is to *reject* the null hypothesis that R^2 is not significantly greater than zero if the computed F-statistic value is *greater* than its critical value.

The computed F-statistic value in our example is given by

$$F = \frac{571.646/1}{93.354/18} = \frac{571.646}{5.186} = 110.229.$$

Since this F is greater than the critical F, we can infer that the R^2-value of .860 is statistically significant—it is unlikely to have occurred because of some idiosyncrasy in the sample data.

Now that we have covered the mechanics of conducting an F-test, let us formally define a couple of other important concepts implicit in the test, concepts that will be helpful in understanding certain other techniques to be covered later. One such concept is the ***mean square***, which is simply a sum-of-squares value divided by its corresponding degrees of freedom. Both the numerator of the F-statistic ($\mathrm{SS_R}/k$) and its denominator ($\mathrm{SS_E}/n - k - 1$) are thus mean squares. Specifically, $\mathrm{SS_R}/k$ is the ***regression mean square*** and can be interpreted as the average sum of the squared variations explained per independent variable. Term $\mathrm{SS_E}/n - k - 1$ is the ***error mean square*** (or ***residual mean square***) and can be roughly interpreted as the average sum of the squared variations (per data point) left unexplained by the regression equation. With these interpretations you can see why a high regression mean square and a low error mean square, and the resulting high F-value, will tend to support an inference that the R^2-value is statistically significant. The mean square values and the corresponding F-value are provided by the computer output (see Exhibit 18.5). These values differ slightly from those we calculated earlier because of rounding.

Mean Square *A sum-of-squares value divided by its corresponding degrees of freedom.*

Regression Mean Square *Regression-sum-of-squares value divided by its degrees of freedom, k.*

Error (Residual) Mean Square *Error (residual)-sum-of-squares value divided by its degrees of freedom, n − k − 1.*

The square root of the error mean square is called the ***standard error of the regression***. Conceptually, it is a sort of standard deviation of the portion of the variation in the dependent variable left unexplained by the independent variable (or variables in a multiple regression equation). The conventional notation used for the standard error associated with a simple regression equation is $s_{Y/X}$. It is given by

$$s_{Y/X} = \sqrt{\frac{\mathrm{SS_E}}{n - k - 1}}.$$

Standard Error of the Regression *An indicator of the dependent-variable variance left unexplained by the independent variable(s) in the equation; it can be viewed as a standard deviation of the prediction errors made by the regression equation.*

The value of the standard error ($s_{Y/X}$) is shown in the computer output (Exhibit 18.5) as 2.277, which is the square root of the error mean square value of 5.186. The concept of standard error is crucial in constructing confidence intervals around the predictions made by a regression equation. It is also relevant for evaluating whether the slope is significantly different from zero, as discussed next.

Testing the Significance of the Slope

An important element of evaluating the usefulness of a regression equation is checking whether the coefficient associated with an independent variable is significantly different from zero. A regression equation based on sample data (namely, $\hat{Y}_i = a + bX_i$) is a crude reflection of an underlying population regression equation, $\hat{Y}_i = \alpha + \beta X_i$, which can be constructed if data from all the population units are available. In other words, the slope b is merely a sample estimate of the population parameter β. Testing the statistical significance of the slope involves testing the following hypotheses:

$$H_o\colon \beta = 0,$$
$$H_a\colon \beta \neq 0.$$

The appropriate test statistic here is a t-statistic defined as

$$t = \frac{b - \beta}{s_b},$$

where s_b is the standard error associated with b and is given by

$$s_b = \sqrt{\frac{s_{Y/X}^2}{(n-1)s_X^2}},$$

where s_X is the standard deviation of the independent variable whose slope is being evaluated. The degrees of freedom for the t-statistic are $n - k - 1$ (d.f. = 18 in our example), and the hypothesis test is two-tailed. The critical values of t from Appendix D for a significance level of .05 are +2.10 and −2.10. Therefore we should reject H_o if $t > 2.10$ or if $t < -2.10$.

Before computing t, we need to determine s_b. The standard deviation of advertising (s_X) turns out to be 4.534. Hence

$$s_b = \sqrt{\frac{(2.277)^2}{(20-1)(4.534)^2}} = .115.$$

This value is shown under STD ERROR B in the computer output in Exhibit 18.5. Thus t is

$$t = \frac{b - \beta}{s_b} = \frac{1.210 - 0}{.115} = 10.522.$$

Since $t > 2.10$, we can reject H_o and conclude that the slope of 1.210 is significantly different from zero.

In simple regression analysis the t-test for checking the significance of b is equivalent to the F-test for checking the significance of R^2. Indeed, the t-statistic value is simply the square root of the F-statistic value associated with R^2 (you can verify this statement in our example). The equivalence of the two tests can also be seen by picturing the nature of the true regression line when β is zero: The line will be perfectly horizontal, implying thereby that R^2 should be zero. Similarly, when the true R^2 is zero, β should also be zero. In multiple regression analysis, however, there is one overall F-test but several t-tests, one for each b-value. Therefore the notion of equivalence between the F-test for R^2 and the t-tests is not very meaningful (this topic is discussed more fully later).

Standardized Regression Coefficient *Value of the slope when data on all the variables in a regression equation are standardized (reduced to a mean of zero and a standard deviation of one).*

One final comment concerning the computer output in Exhibit 18.5: The number under the heading BETA is *not* the value of the true regression line's slope, β. Rather, it is simply the ***standardized regression coefficient***, that is, the b-value when the data on all the variables are standardized (reduced to a mean of zero and a standard deviation of one) before performing the least-squares regression analysis. Standardized regression coefficients are useful in multiple regression analysis because their magnitudes can be compared across independent variables so as to gain insight into the relative influence of each on the dependent variable.

PRACTICAL APPLICATIONS OF REGRESSION EQUATIONS

Thus far we have covered quite a few details about constructing, interpreting, and evaluating regression equations. Let us now pause to look at some practical applications of simple regression equations before we discuss multiple regression equations.

Let us assume that a regression equation is trustworthy—that is, has a high enough and statistically significant R^2-value. (There is no rule telling us what R^2-value can be considered to be high enough; for practical purposes, however, R^2-values of around .7 or higher are usually adequate.) The regression equation can be helpful in two ways.

First, the regression coefficient, or slope, can indicate how sensitive the dependent variable is to changes in the independent variable. To illustrate, suppose the firm marketing Bright detergent currently spends \$2000 on advertising the detergent in a particular market. It is considering a 20% reduction in its advertising budget and wants to know by how much sales revenue might decline in this market as a result. The information the firm desires can be obtained from the slope b of the regression equation we derived:

$$b = 1.210,$$

which, taking into account the units of measurement of X and Y, means that sales revenue will decline by \$1210 for every \$100 reduction in advertising expenditures. Also

$$\text{proposed reduction in advertising expenditures} = 2000 \times .2 = \$400.$$

Therefore

$$\text{anticipated decline in sales revenue} = \frac{1210 \times 400}{100} = \$4840.$$

Second, the regression equation is a forecasting tool for predicting the value of the dependent variable for a given value of the independent variable. For example, suppose the firm marketing Bright wants to forecast sales revenue from a market in which it intends to spend \$750 on advertising the detergent. The needed forecast can be obtained by simply substituting the value of 7.50 for X_i (recall that advertising expenditures are expressed in hundreds of dollars in the regression equation) into the regression equation and solving for $\hat{Y}_i$:

$$\begin{aligned}\hat{Y}_i &= .163 + 1.210\,X_i = .163 + 1.210(7.5)\\ &= .163 + 9.075 = 9.238.\end{aligned}$$

Since sales revenue is expressed in thousands of dollars in the regression equation, the forecast in dollars is given by

$$9.238 \times 1000 = \$9238.$$

Most practical applications of regression analysis revolve around some form of forecasting. Not all applications are for forecasting sales, however. The examples in Research in Use 18.1 demonstrate a wide variety of circumstances in which regression analysis can provide valuable insights.

PRECAUTIONS IN USING REGRESSION ANALYSIS

The wide popularity of regression analysis and the ready access to computer programs capable of performing it offer a strong temptation to researchers to rush into using the technique without ascertaining whether it is appropriate and without giving much thought to its pitfalls. However, potential users of regression analysis must be aware of several key limitations that, if ignored, can result in erroneous inferences. This section will discuss the limitations in the context of simple regression analysis; the same limitations are also applicable to multiple regression analysis (to be discussed in the next section).

First, just as in the case of correlation analysis, regression analysis is only capable of capturing *linear* associations between dependent and independent variables. The regression technique we have covered is not appropriate when the scatter diagram does not display a meaningful linear trend.[9]

Second, a regression equation with a significant R^2-value does not necessarily imply a *cause-and-effect* association between the independent and dependent variables. This point is particularly noteworthy; the convention of labeling one variable "dependent" and the other "independent" may persuade careless researchers into thinking that the former is definitely caused by the latter when the regression results are statistically significant. But the fact of the matter is that a particular direction of causality between two variables must stem from prior knowledge and theoretical considerations, rather than from mathematical manipulation of data on the variables. Thus specification of which is the

---

[9] Modifications of the basic technique are available for dealing with situations involving nonlinearity. For details, see Wonnacott and Wonnacott, *Econometrics*, Chap. 4.

Research in Use 18.1 *Regression Applications*

Illustrative Scenarios Calling for Regression Analysis

Scenario	Possible Dependent Variable	Possible Independent Variable
Curtis is a construction industry lobbyist in an area of the country that is economically depressed. His current charge is to convince local government officials to vote in favor of several tax concessions for the construction industry. He is wondering whether he can generate any concrete evidence to show that increased construction activity (presumably *spurred by the proposed* tax concessions) would greatly benefit the state.	Number of people unemployed or the unemployment rate; data on this variable may be gathered from a sample of areas from around the country.	Number of construction permits issued or number of ongoing construction projects; data on this variable should be gathered from the same sample areas.
Carol, chief librarian in a major university, is eager to increase the number of students borrowing books from the library as well as the number of books borrowed per student. However, she feels that she needs some persuasive evidence to show how increased borrowing of books might benefit them.	Cumulative grade point ratio; data on this variable should be gathered for a sample of students who have borrowed books in the past.	Number of books borrowed; assuming that the library has records of the books borrowed by students, data on this variable can be obtained from those records for the same sample of students.
Jack, an officer in an association in charge of putting together and promoting industrial trade shows, is wondering about the nature and the extent of the impact that the number of exhibitors in a trade show has on trade show attendance.	Number of people visiting a trade show; from the association's past records data on this variable can be obtained for a representative sample of trade shows.	Number of exhibitors in a trade show; here, again, the necessary data can be obtained from past records.

dependent variable and which is the independent variable is done by the *researcher*, not by the regression technique.

To illustrate, consider the second scenario in Research in Use 18.1. Suppose Carol, the librarian, conducts a regression analysis of data on cumulative grade point ratio (dependent variable) and number of books borrowed (independent variable) and obtains a statistically significant R^2-value of .9. Does this result mean high grade point ratios are a result of increased borrowings from the library? The inference does not necessarily follow. Although Carol may be tempted to make such an inference, the following inference is also a plausible explanation for the regression results: Students borrowing a large number of books may have done so *because* they were studious and had high grade point ratios to begin with. In other words, Carol can switch the dependent and the independent variables in the regression equation and still obtain an equally high R^2-value.

Third, a regression equation may not yield a trustworthy prediction of the dependent variable when the value of the independent variable at which the prediction is desired is outside the range of values used in constructing the equation. For example, the range of advertising expenditures used in deriving the regression equation for Bright detergent's sales was $500 to $1900. We take a risk in using this regression equation to predict sales revenue corresponding to an advertising expenditure level of, say, $50 or $2500. We do not know whether the scatter diagram would follow the same linear trend if it were extended at either end by plotting new data points. In fact, even within the original range of independent-variable values, prediction errors will increase as we move away from the sample mean ($\bar{X}$) value in either direction.

Point Estimate
A dependent-variable value predicted by using a regression equation.

Prediction Interval
A confidence interval constructed around a point estimate (for the dependent variable) generated by a regression equation.

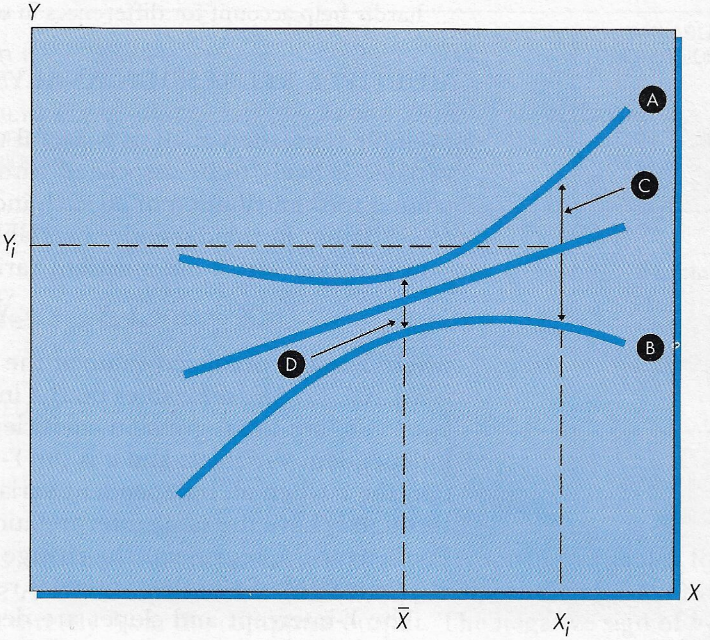
The regression equation's prediction for a given value of the independent variable is merely a ***point estimate***. We can construct a ***prediction interval*** (similar to the 95% confidence intervals discussed in Chapter 12) around each such estimate. A 95% prediction interval, for instance, is the range of values within which we can be 95% confident that the actual value of the dependent variable (corresponding to the given value of the independent variable) will fall. Formulas needed to construct prediction intervals are available in many statistics textbooks.[10] The resulting intervals at various levels of the independent variable form a sort of hourglass shape, as shown in Exhibit 18.7.

Fourth, a regression equation based on relatively few data points cannot be trusted. As we saw earlier, a simple regression equation based on just two data points will have an R^2-value of 1, no matter where those points are. Sample size and, hence, the number of data points available are especially critical in multiple regression analysis, which, as we will

---

[10] For these formulas and the rationale underlying them, see, for instance, Anderson, Sweeney, and Williams, *Introduction to Statistics*, pp. 352–355.

Exhibit 18.7 *Prediction Intervals Corresponding to a Regression Line*



Key

Ⓐ,Ⓑ Upper and lower boundaries of the prediction intervals at various values of X

Ⓒ Prediction interval for the point estimate $\hat{Y}_i$

Ⓓ Smallest prediction interval (occurs at $X = \bar{X}$)

see, involves more than one independent variable. Including too many independent variables when the sample size is small will lead to an artificially high R^2-value. A rule of thumb is to have at least 10 sample units for every independent variable included in the equation.

Fifth, the ranges of data on the dependent and the independent variables can affect the meaningfulness of a regression equation. The ranges for both variables must be sufficiently wide if the regression equation is to be useful. Deriving a meaningful regression line when data on either variable span only a narrow range will be difficult. The need for a sufficient degree of dispersion in the data is succinctly stated by Loether and McTavish:

> Clearly, if there is no variation in the dependent variable, there is nothing to explain. Thus there are relatively few studies of the number of heads humans are born with. Given some variation, we can ask why, and introduce independent variables to help explain. There is also no progress to

> be made in explanation unless independent variables have some variation. If there were no variation in independent variables, then only one prediction would always be made for the dependent variable, which would hardly help account for differences in scores on the dependent variable.[11]

MULTIPLE REGRESSION ANALYSIS

Multiple regression analysis is useful when more than one independent variable is likely to be associated with a dependent variable and ascertaining the contribution of all such independent variables in accounting for variation in the dependent variable is necessary. A multiple regression equation with k independent variables can be written as follows:

$$\hat{Y}_i = a + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki},$$

where $\hat{Y}_i$ is the predicted value of the dependent variable for some unit i; $X_{1i}, X_{2i}, \ldots, X_{ki}$ are values on the independent variables for unit i; $b_1, b_2, \ldots, b_k$ are the regression coefficients (slopes) for the corresponding independent variables; and a is the Y-intercept representing the prediction for Y when all independent variables are set to zero. The slopes b_1 through b_k have the same interpretation as in simple regression analysis. For instance, b_k represents the change in Y per unit change in X_k, assuming the values of all other variables remain the same.

The Y-intercept and slopes are determined by using a least-squares procedure similar to the one we discussed for simple regression analysis. Their values can be readily obtained by using standard computer packages once a data set (like the one in Table 18.7) has been created.

Advertising expenditures for Bright detergent was the only independent variable in the regression equation discussed in the preceding section. But as we saw when we discussed correlation analysis, number of competing brands was also strongly associated with Bright's sales. Therefore examining a regression equation with both advertising and number of competing brands as independent variables may be interesting. The SPSS computer output for the multiple regression analysis of the data set in Table 18.7 is shown in Exhibit 18.8.

Advertising expenditures and number of competing brands are denoted as ADVTG and NCOMP, respectively, in the computer output. Except for the addition of one row corresponding to NCOMP, the format of this computer output is the same as the format of the simple regression analysis output (Exhibit 18.5). Using Y to represent sales (in thousands of dollars), X_1 to represent ADVTG (in hundreds of dollars), and X_2 to represent NCOMP (in actual number of competing detergent brands), we can write the regression equation as

---

[11] Herman J. Loether and Donald G. McTavish, *Descriptive Statistics for Sociologists: An Introduction* (Boston: Allyn and Bacon, 1974), p. 360.

Exhibit 18.8 *SPSS Computer Output for a Multiple Regression Analysis of the Data in Table 18.7*

```
                                    MULTIPLE REGRESSION

DEPENDENT VARIABLE. .    SALES                                                     VARIABLE LIST 1
VARIABLE(S) ENTERED ON STEP NUMBER 1. .    NCOMP                                 REGRESSION LIST 3
                                           ADVTG

MULTIPLE R             0.93421   ANALYSIS OF VARIANCE    DF   SUM OF SQUARES    MEAN SQUARE          F
R SQUARE               0.87274   REGRESSION              2.        580.37308      290.18654   58.29317
ADJUSTED R SQUARE      0.85777   RESIDUAL               17.         84.62692        4.97805
STANDARD ERROR         2.23116

              VARIABLES IN THE EQUATION                         VARIABLES NOT IN THE EQUATION

VARIABLE            B          BETA      STD ERROR B      F       VARIABLE  BETA IN  PARTIAL  TOLERANCE  F
NCOMP       -0.4975703    -0.32847       0.37579      1.753
ADVTG        0.8081323     0.61931       0.32371      6.232
(CONSTANT)    8.854438
ALL VARIABLES ARE IN THE EQUATION
STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.
```

$$\hat{Y} = 8.854 + .808X_1 - .498X_2.$$

In this equation the subscript i is omitted from the variables for simplicity. Also, the values of a (8.854), b_1 (.808), and b_2 ($-.498$) are obtained directly from the computer output. The negative sign of b_2 stems from the inverse relationship between sales of Bright and number of competing brands (see Exhibit 18.2).

The R^2-value for this multiple regression equation is .873. Thus advertising expenditures and number of competing brands *together* account for a little over 87% of the total variation in sales of Bright. The F-statistic value for this R^2 is 58.293, with degrees of freedom $k = 2$ for the numerator and $n - k - 1 = 17$ for the denominator. The critical F-value for a significance level of .05 is 3.59 (from Appendix F). Since the calculated F-value is greater than the critical value, we can infer that the R^2-value is statistically significant and that the regression equation is trustworthy.

A statistically significant R^2-value in multiple regression analysis does not necessarily mean that all regression coefficients (the b's) are significantly different from zero. Each b-value must be checked separately for statistical significance, using a t-test similar to the one for checking the significance of a simple regression coefficient. Since there are two independent variables in our multiple regression equation, we must test two sets of hypotheses:

1. $H_o : \beta_1 = 0$ and $H_a : \beta_1 \neq 0$.
2. $H_o : \beta_2 = 0$ and $H_a : \beta_2 \neq 0$.

The squared value of the t-statistic for each regression coefficient is shown as an F-value in the computer output in Exhibit 18.8 (recall that the t-statistic is the square root of the corresponding F-statistic). The degrees of freedom for the t-statistic are $n - k - 1 = 17$. Assuming a significance level of .05, the critical values for t for a two-tailed hypothesis test are 2.11 and -2.11 (from Appendix D). The actual t-values for b_1 and b_2 can be derived from their corresponding F-values shown in Exhibit 18.8. For b_1

$$t = \sqrt{6.232} = 2.496.$$

For b_2

$$t = -\sqrt{1.753} = -1.324$$

(the negative sign is included because b_2 is negative).

Comparing these t-values with the critical t-values, b_1 is significantly different from zero, but b_2 is not. In other words, according to our multiple regression results, while advertising apparently has a significant association with sales, the number of competing brands does not. This finding is somewhat puzzling because, as we saw earlier, the Pearson correlation coefficient between sales and number of competing brands is quite high ($-.909$) and statistically significant. The explanation for this apparent anomaly lies in the fact that the two independent variables are strongly associated with each other. This association can be seen by scanning the last two columns in the data set shown in Table 18.7. When independent variables in a multiple regression equation are highly correlated among themselves, the problem of ***multicollinearity*** is said to exist. Interpretation of individual regression coefficients is very difficult and can be misleading when multicollinearity is present, as we will see in the next section.

Multicollinearity
A condition wherein two or more independent variables in a multiple regression equation are highly correlated.

MULTICOLLINEARITY'S IMPLICATIONS

Multicollinearity is a frequently encountered problem. Evidence of its presence must be checked for before one interprets a multiple regression equation. Most computer packages will provide a matrix of correlations among all variables as part of the regression analysis output. Exhibit 18.9 contains the matrix of correlations printed by the SPSS program used to derive the multiple regression equation for the sales of Bright.

The high correlation between ADVTG and NCOMP in Exhibit 18.9 implies that interpreting their regression coefficients is risky. Every user of the multiple regression technique must examine such a correlation matrix and exercise caution in interpreting the coefficient (slope) of any independent variable that is strongly associated with one or more other independent variables.

A comprehensive discussion of the rationale underlying the difficulty in interpreting regression coefficients when there is multicollinearity can

Exhibit 18.9 *SPSS Computer Output for a Pairwise Correlation Analysis of the Data in Table 18.7*

	SALES	ADVTG	NCOMP
SALES	1.00000	0.92716	−0.90889
ADVTG	0.92716	1.00000	−0.93721
NCOMP	−0.90889	−0.93721	1.00000

be found elsewhere.[12] However, we can intuitively see why there is difficulty if we reexamine the general interpretation of a regression coefficient: It is the change in the dependent variable per unit change in the independent variable, *assuming all other variables remain the same.* The emphasized words here refer to a condition that cannot be met when there is multicollinearity. That is, a high correlation between two independent variables implies that a change in one will be accompanied simultaneously by a corresponding change in the other. Hence one will have difficulty sorting out the *individual* contributions of the independent variables in accounting for variations in the dependent variable.[13] So interpreting the magnitudes of the regression coefficients as accurate measures of the dependent variable's sensitivity to changes in corresponding independent variables will not be meaningful.

The simple and multiple regression analyses of the data related to Bright detergent yielded two different values for the regression coefficient of advertising expenditures. It was 1.210 in the simple regression analysis (Exhibit 18.5) and .808 in the multiple regression analysis (Exhibit 18.8), despite the fact that the same set of data on sales and advertising was used in both cases. The lower *b*-value in the multiple regression analysis is due to advertising's high correlation with number of competing brands. Neither the regression coefficient of .808 for ADVTG nor the regression coefficient of −.497 for NCOMP have a clearcut interpretation because of the extensive overlap between the two variables. Had the correlation between them been weak, advertising's regression coefficient in the multiple regression analysis would have been close to its original value of 1.210. Moreover, we could have meaningfully interpreted the standardized regression coefficients (listed under the heading BETA on the computer output) for NCOMP and ADVTG. Specifically, *if NCOMP and ADVTG were only weakly correlated,* the relative magnitudes of their standardized regression coefficients (.328 for NCOMP and .619 for ADVTG) would imply that ADVTG was almost twice as important as NCOMP in accounting for the variation in Bright's sales.

---

[12] An excellent treatment of this issue is given in R. B. Darlington, "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin,* 69 (1968), pp. 161–182.

[13] A technique called *partial correlation analysis* is available for examining the association between a dependent and an independent variable after statistically factoring out the effect of other independent variables. For details, see Fred N. Kerlinger and Elazar J. Pedhazur, *Multiple Regression in Behavioral Research* (New York: Holt, Rinehart and Winston, 1973), Chap. 5.

The presence of multicollinearity does not necessarily mean that the regression equation is useless. As long as the R^2-value is high and statistically significant, the equation can still be used for predicting Y if one is given values for all the independent variables. One caveat, however, is that the relative values of the independent variables for which the prediction is desired must be consistent with the general pattern of values in the data set used to derive the multiple regression equation.

For instance, consider the multiple regression equation derived for the sales of Bright. The input data on advertising and number of competing brands (Table 18.7) are never *both* high or low. For this reason the two variables have a strong negative correlation. If we want to predict Bright's sales for a pair of new values of the independent variables, one of them must be high and the other low for the prediction to be meaningful. Thus, for example, the equation's prediction of sales when advertising is $650 and number of competing brands is 2 is unlikely to be trustworthy—both values are low vis-à-vis the range of values in the data set.

Multicollinearity is not the only potential limitation of multiple regression analysis. The precautions we discussed under simple regression analysis apply to multiple regression analysis as well, and they must be heeded to avoid making misleading inferences. These precautions include (1) bearing in mind that regression analysis only captures the extent of linear relationship between dependent and independent variables; (2) not inferring *cause-and-effect* associations solely from statistically significant regression results; (3) exercising caution in predicting the dependent variable for independent variable values that fall outside the ranges used in constructing the regression equation; (4) ensuring that there are enough data points (a rule of thumb is to have at least 10 data points for each independent variable); and (5) ensuring that the ranges of data on all variables are sufficiently wide.

A few final remarks about regression analysis: Despite the rather extensive treatment of the technique in this chapter, we have only covered the fundamentals necessary to give you a basic understanding of it, including its capabilities and limitations. There is much more to regression analysis than what a textbook of this nature can possibly cover. Not covered here are topics such as transforming data on variables to make them satisfy the assumptions of regression analysis. (One topic—that of including categorical independent variables in a regression analysis by transforming them into dummy variables—will be covered in the next chapter when we discuss analysis of variance.) While such topics are not relevant for all regression applications, students wishing to gain a comprehensive understanding of regression analysis will benefit from additional reading.[14]

SUMMARY

This chapter deals with procedures for ascertaining and interpreting associations between variables. A frequently used technique for examining associations when data on the variables are nonmetric (nominal or ordinal) is the chi-square contingency test. It involves comparing the observed cell frequencies in a two-way table (called a contingency table) with corresponding expected cell frequencies computed under the null hypothesis of no association between the variables. Certain minimum cell frequency requirements must be met for the chi-square contingency test to be meaningful. Also, this test by itself only indicates whether or not two variables are significantly related. When this test leads to the rejection of the null hypothesis of independence, some idea about the degree of association between the variables can be obtained by computing the contingency coefficient C or Cramer's V-statistic. However, these measures of association are somewhat arbitrary. Their values are sensitive to how the variable categories are defined and how many rows and columns the contingency table has.

The Spearman correlation coefficient r_s is a better measure of association when the data on variables are ordinal and are in the form of ranks of the sample units on each variable. This coefficient will range from $+1$ to -1, with a value of zero signifying no correlation. Its statistical significance can be verified by conducting a t-test.

The Pearson correlation coefficient is a widely used measure of association between metric (interval or ratio) variables. It is a number ranging from $+1$ to -1 and summarizes the degree of compactness as well as the direction of a scatter diagram, which is a plot of the data on the variables. Whether this correlation coefficient is significantly different from zero can be determined easily, given its value, the sample size, and a desired significance level. A limitation of the Pearson correlation coefficient is that it can only capture the degree of linear association between variables. Furthermore, evidence of a strong correlation by itself does not constitute evidence of causality in any particular direction.

Simple regression analysis is similar to correlation analysis except that it generates a mathematical relationship (called the regression equation) between one variable designated as the dependent variable (Y) and another designated as the independent variable (X). An independent variable is typically assumed to influence the dependent variable in some fashion. The regression equation, $\hat{Y}_i = a + bX_i$, is derived by fitting a straight line through the scatter diagram so as to minimize the error sum of squares (SS_E). This procedure is called the least-squares technique and is available from several computer packages. The constant a is called the Y-intercept, and the regression coefficient b is called the slope. The slope has an important interpretation: It is the change in the predicted value of Y per unit change in X, when all other variables likely to influence X remain the same.

---

[14] Several books on regression analysis are available, such as Kerlinger and Pedhazur, *Multiple Regression*, and Johnston, *Econometric Methods*.

Several indicators are available to check the usefulness of a regression equation. One that is commonly examined is the coefficient of determination R^2. The value of R^2 can range from 0 to 1 and is a measure of the proportion of the total variance in Y that is accounted for by predictions made by the regression equation. The higher the R^2-value, the better the regression equation is. The statistical significance of R^2 can be checked by using an F-test. The square root of R^2 is called the multiple correlation coefficient. In bivariate regression analysis the multiple correlation coefficient is the same as the Pearson correlation coefficient. A related measure of goodness of a regression equation is the standard error of the regression, $s_{Y/X}$, which is a sort of average variation in Y left unexplained by the regression equation. The smaller the value of $s_{Y/X}$, the better the regression equation is.

Another way to evaluate a regression equation is to see whether the slope is significantly different from zero. A t-test is employed for this purpose. The F-test for the significance of R^2 and the t-test for the significance of b are equivalent in simple regression analysis.

Two practical applications of simple regression analysis are to ascertain changes in the dependent variable when the independent variable changes by a specified amount, and to predict the value of the dependent variable for a given value of the independent variable. In interpreting and using the regression equation, however, researchers must keep several potential pitfalls in mind. Regression analysis by itself cannot establish causality between variables. Regression equations based on nonlinear scatter diagrams, on a limited number of data points, or on data spanning relatively narrow ranges are unlikely to be meaningful. Also, making predictions of the dependent variables for values of the independent variable that fall outside the range of data used in deriving the regression equation is risky.

Multiple regression analysis is a logical extension of simple regression analysis and involves two or more independent variables. The analytical procedures and the interpretation of the results are quite similar in both cases. The limitations of simple regression analysis apply to multiple regression analysis as well. A key difference between the two techniques is that statistical significance of the R^2-value of a multiple regression equation (based on an F-test) does not automatically mean that all the regression coefficients (b-values) are significantly different from zero. Each regression coefficient must be individually evaluated by using a

t-test. Furthermore, a potentially serious problem in multiple regression analysis is multicollinearity—that is, the presence of strong correlations among independent variables. When multicollinearity exists, caution must be exercised in interpreting the values of individual regression coefficients.

KEY TERMS

Chi-square contingency test
Contingency table
Contingency coefficient
Cramer's V-statistic
Spearman correlation coefficient
Pearson correlation coefficient (product moment correlation)
Scatter diagram
Simple regression analysis
Multiple regression analysis
Independent (explanatory or predictor) variable
Dependent (criterion) variable
Least squares approach
Y-intercept
Slope
Coefficient of determination (R^2)
Total sum of squares
Regression sum of squares
Error (residual) sum of squares
Multiple correlation coefficient
Mean square
Regression mean square
Error (residual) mean square
Standard error of the *regression*
Standardized-regression coefficient
Point estimate
Prediction interval
Multicollinearity

QUESTIONS

1. From a survey of 1000 families the following two-way table was constructed:

	Owned Product X	
Income Level	**Yes**	**No**
Hi	200	150
Med	175	100
Lo	125	250
	500	500

Is there a statistically significant relationship between income level and ownership of product X, at the 95% confidence level?

2. A random sample of 500 consumers in a certain county consisted of 300 males and 200 females. Of the males in the sample 150 were smokers, and of the females 75 were smokers. On the basis of this evidence, is there a statistically significant relationship (at the 95% confidence level) between the sex of consumers in this county and whether or not they smoke?

3. A computer retailer in a large city wanted to see whether the purchase of Apple computers by families was associated with whether or not families had children living at home. A random survey of 900 families in the city showed that 300 of them had children living at home. The survey also showed 40% of the families with children owned Apple computers while 30% of the families without children owned Apple computers. Is there a statistically significant association between children living at home and Apple computer ownership (assume a significance level of .01)?

4. What are the potential limitations of the chi-square contingency test?

5. Compute and interpret the contingency coefficient and Cramer's *V*-statistic for the problems in Questions 1, 2, and 3. Is the contingency coefficient more meaningful than Cramer's *V*-statistic (or vice versa) in each case? Why or why not?

6. What major caveat should one bear in mind in computing and interpreting numbers apparently representing the *degree* of association between categorical variables that have been cross-tabulated to produce a contingency table?

7. Ace Chemicals, Inc. (ACI), selected 15 of its customer firms at random and rank-ordered them from highest (rank 1) to lowest (rank 15) on the basis of their volumes of purchases from ACI. It then rank-ordered these same 15 firms according to how formalized the firms' purchasing operations were (rank for most formalized = 1 and rank for least formalized = 15). For the firms ranked 1 through 15 in terms of purchase volumes, the following ranks reflect the degree of formalized purchasing operations, respectively: 3, 4, 2, 1, 5, 8, 13, 7, 6, 11, 9, 10, 15, 12, 14. In other words, the firm with the largest purchase volume ranked third with respect to formalized purchasing operations, the firm with the second largest purchase volume ranked fourth with respect to formalized purchasing operations, and so on. Compute the Spearman correlation coefficient between customer purchase volume from ACI and formalization of customers' purchasing operations. (If you have access to a computer, use it to derive the Spearman correlation coefficient with the aid of a canned computer program.) How do you interpret this correlation coefficient?

8. Is the correlation coefficient you computed in Question 7 statistically significant? Why or why not?

9. The sales manager of a firm administered a standard, multi-item job satisfaction scale to a sample of the firm's sales force. The manager then correlated the satisfaction scores with the ages of the salespeople in the sample. The Pearson correlation coefficient between satisfaction and age turned out to be .08. On the basis of this evidence the sales manager came to the following conclusions: "A salesperson's age has hardly anything to do with his or her satisfaction. Furthermore, as salespeople get older, they continue to have the same average levels of job satisfaction." Would you agree or disagree with the sales manager's conclusions? Explain your answer.

10. Compute the Pearson correlation coefficient for the data in Question 7 (using a canned computer program, if possible). How does this coefficient compare with the Spearman correlation coefficient computed earlier? From this comparison, what inference can you make about marketing researchers' often assuming ordinal data to have interval properties?

11. What are the key differences between simple regression and correlation analyses?

12. Clearly summarize, in a sentence or two, what the least-squares approach to regression analysis does.

13. Exhibit 18.10 shows the SPSS output for the simple regression analysis of the data on sales (SALES, expressed in thousands of dollars) and number of competing detergents (NCOMP) given in Table 18.7. Notice that the format of Exhibit 18.10 is identical to that of Exhibit 18.5. Use the output of Exhibit 18.10 to answer the following questions.

a. Identify the Y-intercept and slope values, and interpret those values.

b. Write the regression equation linking predicted sales of Bright and the number of competing detergent brands.

c. What is the R^2-value associated with the regression equation? Is it statistically significant? Why or why not?

d. Verify that $R^2 = SS_R/SS_T$, and give a verbal interpretation of the R^2-value.

e. Circle the number representing the Pearson correlation coefficient between the two variables.

f. Verify that the square root of the error mean square is the standard error of the regression.

Exhibit 18.10 *SPSS Computer Output for a Simple Regression Analysis of Sales and Number of Competing Brands*

```
                                 MULTIPLE REGRESSION

DEPENDENT VARIABLE. .     SALES                                                 VARIABLE LIST 1
VARIABLE(S) ENTERED ON STEP NUMBER 1. .     NCOMP                              REGRESSION LIST 2

MULTIPLE R               0.90889   ANALYSIS OF VARIANCE    DF   SUM OF SQUARES   MEAN SQUARE        F
R SQUARE                 0.82809   REGRESSION               1.       549.34783     549.34783   85.50000
ADJUSTED R SQUARE        0.81643   RESIDUAL                18.       115.65217       6.42512
STANDARD ERROR           2.53478

               VARIABLES IN THE EQUATION                          VARIABLES NOT IN THE EQUATION

VARIABLE            B           BETA      STD ERROR B      F       VARIABLE  BETA IN  PARTIAL  TOLERANCE  F
NCOMP         -1.376812     -0.90889        0.14890     85.500
(CONSTANT)    25.37681

MAXIMUM STEP REACHED
STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.
```

g. Is the slope of the regression equation statistically significant? Why or why not?

h. Use the implied regression equation to predict sales for Bright in a market where it will have 19 competing detergent brands. Is this prediction meaningful? Explain your answer.

i. Bright currently has five competing detergent brands in a certain market area. The firm expects five additional competing brands to enter the market in the near future. What impact will the new detergent brands have on Bright's monthly dollar sales in this market area? What precautions should you take in interpreting the estimated impact?

14. What is multicollinearity? What implications does the presence of multicollinearity have for the interpretation and use of a regression equation?

15. You are given the following results for a regression analysis of a dependent variable (Y) and two independent variables (X_1 and X_2) from a sample of 30 observations:

Means	$\overline{Y} = 3000$	$\overline{X}_1 = 100$	$\overline{X}_2 = 150$
Standard deviations	$S_Y = 25$	$S_{X_1} = 30$	$S_{X_2} = 20$
Correlation coefficients	$r_{YX_1} = .75$	$r_{YX_2} = .95$	$r_{X_1X_2} = .8$
Estimated regression equation	$\hat{Y} = 1000 + 5X_1 + 10X_2$		
$R^2 = .85$	F-value associated with $R^2 = 5.78$		

a. Is the R^2-value statistically significant? Why or why not?

b. Can you meaningfully determine the change in Y when X_1 is increased by one unit? If you can, what is the change in Y? If you cannot, why not?

c. What proportion of the variance of Y is *not* explained by the regression equation? What does this result say about how good the regression equation is?

Some Technical Details About Correlation and Regression Analysis

DERIVATION AND INTERPRETATION OF PEARSON CORRELATION COEFFICIENT

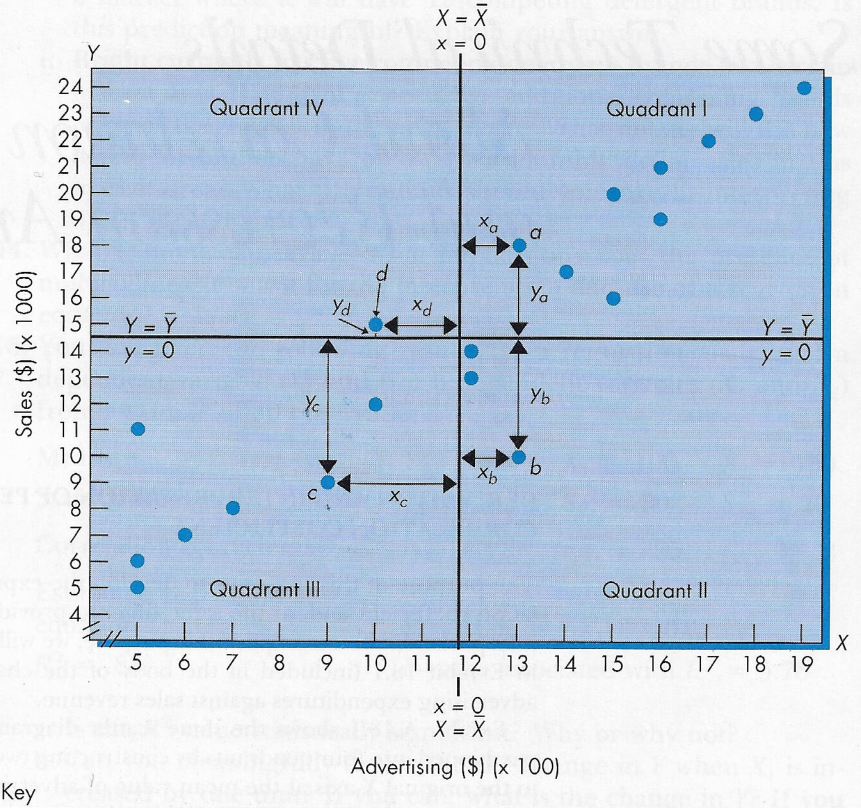
The purpose of this section is to develop the expression for the Pearson correlation coefficient and, at the same time, to provide the rationale underlying its interpretation. To accomplish this purpose, we will start with the scatter diagram in Exhibit 18.1 (included in the body of the chapter), which shows a plot of advertising expenditures against sales revenue.

Exhibit A.18.1 shows the same scatter diagram as in Exhibit 18.1, but it is partitioned into four quadrants by constructing two new axes, one perpendicular to the original X-axis at the mean value of advertising expenditures ($\bar{X}$) and the other perpendicular to the original Y-axis at the mean value of sales ($\bar{Y}$). For each data point in Exhibit A.18.1, the uppercase coordinates (X, Y) refer to the raw data, and the lowercase coordinates (x, y) refer to deviations of the raw data from their respective mean values. The coordinates (x, y) are central to determining the Pearson correlation coefficient.

Consider the four illustrative data points a, b, c, and d in quadrants I, II, III, and IV, respectively. What can we say about the product xy for each of these data points? For data point a both x_a and y_a are positive. Therefore $x_a y_a$ is also positive. Indeed, the xy-product is positive for all data points in quadrant I. Now examine data point c in quadrant III. Since both x_c and y_c are negative, their product ($x_c y_c$) is positive. Thus the product of the deviations x and y is positive for all data points in quadrants I and III—the points that, in general, are most consistent with a linear, upward-sloping trend, signifying a positive association between the variables.

Data points in quadrants II and IV are off-diagonal relative to the linear, upward-sloping trend. Intuitively, if too many data points fall in quadrants II and IV, the evidence of a direct positive association between the variables will be weakened. Not accidentally, the product xy is negative for all data points in these

Exhibit A.18.1 *Plot of Deviations of Sales and Advertising Data from Their Respective Means*



Key

$\bar{X}$ = mean advertising expenditure ($1185)

$\bar{Y}$ = mean sales ($14,500)

$x = X - \bar{X}$

$y = Y - \bar{Y}$

two quadrants. For instance, for data point *b*, x_b is positive but y_b is negative, and for data point *d*, x_d is negative but y_d is positive. Thus the product of the deviations will be negative in both cases.

Keeping in mind the foregoing discussion about the nature of the data points in the four quadrants, examine the following quantity, which is called the *covariance* of *X* and *Y* and denoted as Cov(*X*, *Y*):

$$\mathrm{Cov}(X, Y) = \frac{\sum_{i=1}^{n} x_i y_i}{n - 1},$$

where n is the sample size (total number of data points). The Cov(X, Y) is an average measure of the degree to which X and Y move together. Clearly, the larger the number of data points in quadrants I and III, the higher the value of Cov(X, Y) will be. In contrast, the larger the number of data points in quadrants II and IV, the smaller the value of Cov(X, Y) will be. Alternatively, a large positive value for Cov(X, Y) will signify a strong direct association between X and Y, and a large negative value for Cov(X, Y) will signify a strong inverse association between X and Y.

Although Cov(X, Y) is a quantitative measure of the degree of association between variables, one problem in interpreting it is that it can be arbitrarily increased by merely changing the units of measurement for the variables. For example, by using cents instead of dollars for both sales and advertising in the expression for Cov(X, Y), the value of the covariance between the two can be increased by a factor of 10,000. Hence for a clear interpretation of covariance, its value must be made independent of the units in which the variables are measured. This independence can be accomplished by standardizing the variables. The resulting standardized covariance is the Pearson correlation coefficient, given by the following expression:

$$r_{XY} = \frac{\sum_{i=1}^{n} x_i y_i}{(n - 1)s_X s_Y},$$

where s_X and s_Y are the standard deviations of the two variables. The value of r_{XY} can range from $+1$ to -1.

LEAST SQUARES REGRESSION ANALYSIS

This section briefly reviews how the least-squares approach derives the best-fitting regression line for a given scatter diagram. The rationale underlying the least squares approach stems from the implied primary purpose of a regression equation, namely, to predict values of the dependent variable (Y) for given values of the independent variable (X), *with errors as small as possible*.

Consider the upward-sloping line in Exhibit A.18.2, which contains the same scatter diagram as in Exhibit 18.1 presented in the chapter. For any given value of X, say X_i, we can predict a Y-value, say, $\hat{Y}_i$, using the line in Exhibit A.18.2. However, $\hat{Y}_i$ will not necessarily coincide with the actual value Y_i corresponding to X_i. Thus using this line to predict a Y-value for a given X_i-value will lead to a ***prediction error*** (e_i), defined as

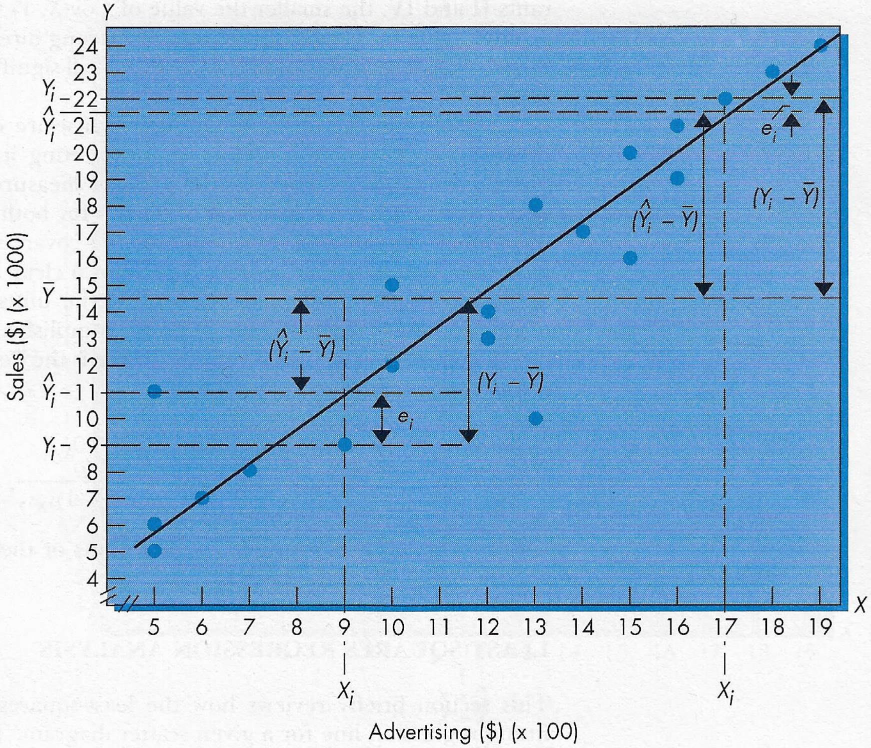
Prediction Error
The difference between the actual value of a dependent variable and its estimated value obtained by using a regression equation.

$$e_i = Y_i - \hat{Y}_i.$$

Illustrative prediction errors pertaining to two data points are shown in Exhibit A.18.2. The quantities ($\hat{Y}_i - \overline{Y}$) and ($Y_i - \overline{Y}$) in this exhibit are discussed in the next section of the appendix. Notice that prediction errors corresponding to data points above the line will be positive, and those corresponding to data points below the line will be negative. Furthermore, different straight lines, such as the ones in Exhibit 18.4 presented in the chapter, will result in different sets of prediction errors.

What the least squares approach does is pick the line that *minimizes the sum of*

Exhibit A.18.2 *Prediction Errors and Explained Variation Associated with a Regression Line*



the squared prediction errors across all data points. In other words, the values of the constant a and b in the equation representing this line (i.e., $Y = a + bX$) are chosen so as to make the quantity Σe_i^2 as small as possible. The mathematics involved in deriving the optimum values of a and b is complex. But the resulting formulas for computing the constants are

$$b = \frac{\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right) \Big/ n}{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n},$$

$$a = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i - b\sum_{i=1}^{n} X_i\right).$$

Once the values of a and b are known, they can be substituted into the regression equation, $\hat{Y}_i = a + bX_i$, where $\hat{Y}_i$ is the predicted dependent-variable value for a given value of X_i for the independent variable.

PARTITIONING THE TOTAL SUM OF SQUARES

Take another look at Exhibit A.18.2. How do the predictions made with the aid of the regression line (i.e., the $\hat{Y}_i$'s) compare with the best prediction that could be made in the absence of the regression line (i.e., $\overline{Y}$)? Clearly, predictions based on the regression line are generally much closer to the actual Y-values. Although the regression line's predictions are not perfect, they are able to account for a substantial part of the total deviation in Y. As demonstrated by Exhibit A.18.2, the following relationship holds for any given value X_i of the independent variable:

$$\begin{aligned}(Y_i - \overline{Y}) &= (\hat{Y}_i - \overline{Y}) + (Y_i - \hat{Y}_i)\\ &= (\hat{Y}_i - \overline{Y}) + e_i.\end{aligned}$$

This relationship, written in words, is

$$\begin{aligned}\text{total deviation} &= \text{deviation explained by regression line}\\ &\quad + \text{deviation unexplained by regression line.}\end{aligned}$$

Using this relationship as a starting point, we can show that

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} e_i^2.$$

Each term in this expression represents the summation of a particular type of squared deviation and is hence called a *sum of squares*:

$\sum_{i=1}^{n} (Y_i - \overline{Y})^2$	*Total* sum of squares; a measure of the total *variance* in the dependent variable; denoted as SS_T
$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$	*Regression* sum of squares; represents the portion of the variance explained by the regression equation; denoted as SS_R
$\sum_{i=1}^{n} e_i^2$	*Error* (or sometimes residual) sum of squares; represents the portion of the variance unexplained by the regression equation; denoted as SS_E

The procedure involved in deriving the relationship $SS_T = SS_R + SS_E$ is called *partitioning the total sum of squares*. It is fundamental to regression analysis and to a technique called analysis of variance, to be covered in the next chapter.

CHAPTER 19

Overview of Other Multivariate Techniques

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Distinguish between dependence and interdependence techniques.*
- *Suggest an appropriate ANOVA procedure for analyzing data from a marketing experiment.*
- *Point out the unique, distinguishing features of completely randomized ANOVA, randomized-block ANOVA, Latin-square ANOVA, factorial ANOVA, and ANCOVA.*
- *Describe and illustrate interaction effect in a factorial ANOVA.*
- *State the two key purposes of discriminant analysis.*
- *Interpret, evaluate, and use a discriminant function.*
- *Discuss the purpose and potential applications of factor analysis.*
- *Interpret a factor-loading matrix.*
- *Distinguish between cluster analysis and discriminant analysis.*
- *State the potential uses of multidimensional scaling and point out its key limitations.*
- *State the purpose of conjoint analysis and use the results from such an analysis.*

The preceding chapter discussed a variety of techniques for examining associations, including regression analysis, which is a very widely used multivariate procedure. However, there are many other multivariate techniques, some of which are relatively new and becoming increasingly popular. The purpose of this chapter is to introduce you to several of them and to provide an intuitive understanding of each.

Dependence Technique
A multivariate analysis procedure in which one variable is designated as the dependent variable and the rest are treated as independent variables (e.g., regression analysis, ANOVA, discriminant analysis).

Interdependence Technique
A multivariate analysis procedure in which there are no dependent- and independent-variable designations; all variables are treated equally in a search for underlying patterns of relationships (e.g., factor analysis, cluster analysis).

Multivariate techniques are broadly classified into dependence and interdependence techniques. A ***dependence technique*** is a procedure in which one variable is designated as the dependent variable and the rest are treated as independent variables. Regression analysis is an example of a dependence technique. An ***interdependence technique*** is a procedure in which there are no dependent- and independent-variable designations; rather, all variables are treated equally in a search for underlying patterns of relationships.

Table 19.1 provides an overview of commonly used dependence and interdependence techniques. The techniques differ in terms of the measurement level of the input data and/or in terms of the purpose of the analysis. Of these techniques, regression analysis was discussed in Chapter 18. The remaining techniques will be discussed in this chapter. Since these techniques are rather complex, our coverage of them will be rudimentary. There is more to the techniques—in terms of sophisticated extensions and analytical nuances—than will be discussed here.

ANALYSIS OF VARIANCE

Analysis of variance (ANOVA) is closely related to regression analysis in that its purpose is also to examine the relationship between dependent

Table 19.1 *Overview of Multivariate Techniques*

Type of Technique	Usual Form of the Input Data	Primary Purpose of the Technique
Dependence Techniques		
Regression analysis	Dependent variable, metric Independent variable(s), metric	To ascertain the relative importance of independent variable(s) in explaining variation in the dependent variable; also to predict dependent-variable values for given values of the independent variable(s)
Analysis of variance (ANOVA)	Dependent variable, metric Independent variable(s), nonmetric	To see whether different levels (treatments) of independent variable(s) have significantly different impacts on the dependent variable
Discriminant analysis	Dependent variable, nonmetric Independent variable(s), metric	To identify independent variables that are critical in distinguishing between subsamples defined by the dependent-variable categories; also to aid in classifying new units into one of the subsample categories
Interdependence Techniques		
Factor analysis	Metric	To reduce data on a large number of variables into a relatively small set of factors; also to identify key constructs underlying the original set of measured variables
Cluster analysis	Metric	To identify natural clusters of objects on the basis of similarities of the objects on a variety of characteristics
Multidimensional scaling	Nonmetric (similarity ranks based on comparisons of actual objects)	To identify key dimensions underlying respondent evaluations of products, brands, stores, etc.; also to determine the relative positions of the objects in multidimensional space
Conjoint analysis	Nonmetric (preference ranks based on comparisons of hypothetical stimuli formed by systematically varying selected attribute levels)	To derive utility values that respondents implicitly assign to various levels of key attributes used in evaluating objects; the utility values themselves aid in ascertaining the relative importance of the attributes and the potential attractiveness of descriptive profiles defined by different combinations of attributes

Analysis of Variance (ANOVA) *A technique for detecting relationships between a metric-scaled dependent variable and one or more categorical (nominal or ordinal) independent variables. ANOVA procedures are often used to analyze data from experimental studies.*

Treatment *Each specific level at which an independent variable is set in an ANOVA procedure.*

and independent variables. There is, however, one basic difference between the two techniques: While both dependent and independent variables are metric (i.e., interval or ratio) in regression analysis, only the dependent variable is metric in analysis of variance; the independent variable is categorical (i.e., nominal or ordinal). ANOVA is therefore especially appropriate in situations where the independent variable is set at certain specific levels (called ***treatments*** in an ANOVA context) and metric measurements of the dependent variable are obtained at each of those levels. Not surprisingly, ANOVA procedures are often used to analyze data from marketing experiments employing designs such as the ones covered in Chapter 9. For instance, ANOVA would be an appropriate analysis technique in an experiment in which different groups of stores are exposed to various amounts of in-store promotion for a brand and are measured on some criterion variable (e.g., sales of the brand). Let us look at a specific example to understand what ANOVA does.

Example. A grocery store chain wanted to evaluate the price sensitivity of sales in its store brand of vegetable oil. Twenty-four of the chain's stores in a certain region were chosen at random to participate in an experimental study involving the following three treatments:

1. Store brand sold at the regular price.

2. Store brand sold at 50¢ off the regular price.

3. Store brand sold at 75¢ off the regular price.

Eight of the 24 stores were randomly assigned to the first treatment, another 8 were randomly assigned to the second treatment, and the rest were assigned to the third treatment. Sales of the store brand of vegetable oil were monitored for a week in each store. The data thus gathered are summarized in Table 19.2.

The experimental design implied in this study is the *completely randomized design* described in Chapter 9. Using X's to denote treatments and O's to denote measurements, we can symbolically represent this experimental design as follows:

$$\begin{array}{lll} EG_1(R) & X_1 & O_1 \\ EG_2(R) & X_2 & O_2 \\ EG_3(R) & X_3 & O_3 \end{array}$$

Price is the independent variable (nonmetric, with three categories) in Table 19.2. Unit sales, which is ratio-scaled, is the dependent variable. The basic question now is whether the cents-off treatments had any significant impact on sales. One way of checking this question is to look for evidence of significant differences in mean sales between *pairs* of treatment groups, using the t-test procedure discussed in Chapter 17. A problem with this approach is that a separate t-test is required for each possible pair of treatments. We would have to conduct three t-tests in our example, a potentially cumbersome, and statistically inaccurate, pro-

Table 19.2 *Unit Sales Data Under Different Pricing Treatments*

Treatment	Regular Price ($j = 1$)	50¢ Off ($j = 2$)	75¢ Off ($j = 3$)
Unit sales in each store (Y_{ij})	37	46	46
	38	43	49
	40	43	48
	40	45	48
	38	45	47
	38	43	48
	40	44	49
	39	44	49
Number of stores (n_j)	8	8	8
Mean sales ($\overline{Y}_{.j}$)	38.75	44.13	48.00

cedure. The task would be even more laborious when there are more treatment levels. ANOVA is a more appropriate approach to use when multiple sample means are to be compared. It provides a one-shot, global test for detecting significant differences between treatment group means.

When there are k treatment groups or samples, ANOVA can aid in testing the following hypotheses:

$$H_o{:}\mu_1 = \mu_2 = \cdots = \mu_k.$$

H_a: At least one μ is different from one or more of the others.

Here the μ's are means of the dependent variable in the populations represented by the corresponding treatment groups. The test statistic in this case is an F-statistic based on the concept of partitioning the total sum of squares, introduced while discussing regression analysis in Chapter 18. The following general notation is helpful in developing an understanding of what the F-statistic in ANOVA represents:

Y_{ij} = dependent variable value for sample unit i in treatment group j;

$\overline{Y}_{.j}$ = mean value for treatment group j;

$\overline{Y}_{..}$ = grand mean value (across all units in all treatment groups);

n_j = number of sample units in treatment group j.

The total deviation of any sample unit's value from the grand mean is $(Y_{ij} - \overline{Y}_{..})$. This total deviation can be written as

$$(Y_{ij} - \overline{Y}_{..}) = (Y_{ij} - \overline{Y}_{.j}) + (\overline{Y}_{.j} - \overline{Y}_{..}).$$

With the above expression as a starting point, it can be shown that

$$(Y_{ij} - \overline{Y}_{..})^2 = (Y_{ij} - \overline{Y}_{.j})^2 + (\overline{Y}_{.j} - \overline{Y}_{..})^2. \tag{19.1}$$

Summing Eq. (19.1) across all units in any treatment group j yields

$$\sum_{i=1}^{nj} (Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{nj} (Y_{ij} - \overline{Y}_{.j})^2 + n_j(\overline{Y}_{.j} - \overline{Y}_{..})^2. \tag{19.2}$$

Now summing Eq. (19.2) across all treatments yields

$$\sum_{j=1}^{k} \sum_{i=1}^{nj} (Y_{ij} - \overline{Y}_{..})^2 = \sum_{j=1}^{k} \sum_{i=1}^{nj} (Y_{ij} - \overline{Y}_{.j})^2 + \sum_{j=1}^{k} n_j(\overline{Y}_{.j} - \overline{Y}_{..})^2. \tag{19.3}$$

Total Sum of Squares *Quantity obtained by adding across all data points the squared deviation of each actual value of the dependent variable from its mean value.*

Within Sum of Squares *The portion of the total sum of squares that occurs solely due to variation within the treatment groups.*

The left-hand term in Eq. (19.3) is the ***total sum of squares*** (SS_T) representing the variation across all units and all treatments. The first term on the right-hand side is called the ***within sum of squares*** (SS_W). It is a measure of variation in dependent-variable values within the treatment groups. The second term on the right-hand side is called ***between sum of squares***, or sum of squares attributable to the *treatments* (SS_{TR}). Term SS_{TR} is a weighted measure of variation in values between treatment groups.

Between Sum of Squares *The portion of the total sum of squares that occurs due to differences between the treatment levels.*

It follows that $SS_T = SS_{TR} + SS_W$. Notice the similarity between this expression and the expression $SS_T = SS_R + SS_E$ that we derived in the regression analysis discussion. The treatment sum of squares (SS_{TR}) is analogous to the explained variation (SS_R) under regression analysis, and the within sum of squares (SS_W) is analogous to the residual (or unexplained) variation (SS_E). Intuitively, if the experimental treatments (independent variable) had little impact on the Y-values (dependent variable), we would expect SS_{TR} to be small. For a given value of SS_T, a small SS_{TR} would imply a large SS_W. Just as in regression analysis, a comparison of SS_{TR} and SS_W, taking into account their appropriate degrees of freedom, is the basis for the F-statistic in ANOVA. When there are k treatments, the degrees of freedom associated with SS_{TR} and SS_W are $(k - 1)$ and $(n - k)$, respectively, where n is the total number of units across all treatments. Therefore

$$F = \frac{SS_{TR}/k - 1}{SS_W/n - k}.$$

The numerator and denominator of F are mean squares corresponding to the variation owing to the treatments and chance variation owing to sampling, respectively. To see whether the experimental treatments had a significant impact on the dependent measure, the computed value of F is checked against an appropriate critical value from the F table. Hypothesis H_o is rejected if the computed value exceeds the critical value.

Returning to the data in Table 19.2, let us conduct an ANOVA to see whether at least one of the treatment means (shown in the last row of the table) is significantly different from one or more others. Although

Exhibit 19.1 *SPSS Computer Output for ANOVA Analysis of the Data in Table 19.2*

```
                                ANALYSIS OF VARIANCE
SALES BY TREAT
                                SUM OF                MEAN                  SIGNIF
SOURCE OF VARIATION             SQUARES       DF      SQUARE        F       OF F
MAIN EFFECTS                    345.250        2      172.625     137.447   0.000
   TREAT                        345.250        2      172.625     137.447   0.000
EXPLAINED                       345.250        2      172.625     137.447   0.000
RESIDUAL                         26.375       21        1.256
TOTAL                           371.625       23       16.158
   24 CASES WERE PROCESSED.
    0 CASES (0.0 PCT) WERE MISSING.
```

we can compute the necessary *F*-value by hand, using the formulas just derived, it is more convenient to use one of the readily available computer packages for this purpose. Given the raw dependent-variable values for each treatment (i.e., the Y_{ij}-values), a computer program can perform ANOVA efficiently and provide a rich variety of output information. The computer output resulting from analyzing the raw data in Table 19.2 by using the SPSS ANOVA program is shown in Exhibit 19.1.

The names SALES and TREAT were used for the dependent and independent variables in the computer program (they are shown in the output just below the heading). A tabular array such as the one in Exhibit 19.1, showing various sums of squares, degrees of freedom, mean squares, *F*-values, and their corresponding significance levels, is usually called an ANOVA table. Its format is similar to the formats we saw in the regression analysis computer outputs (Exhibits 18.5 and 18.8).

The number 137.447 shown under F in the row labeled TREAT in Exhibit 19.1 is the computed *F*-value we are looking for. It is the ratio of the mean square value of 172.625 associated with TREAT (that is, $\text{SS}_{\text{TR}}/k - 1$) and the mean square value of 1.256 associated with RESIDUAL (that is, $\text{SS}_{\text{W}}/n - k$). The critical *F*-value with 2 numerator degrees of freedom and 21 denominator degrees of freedom at the traditional significance level of .05 is only 3.47 (from Appendix F). We should therefore reject H_o and conclude that at least one of the treatment means is significantly different from one or more others. Indeed, in this case we need not even have determined the critical value of *F* to arrive at the above conclusion, since the ANOVA output shows the actual significance level of the *F*-test—the value 0.000 under the column SIGNIF OF F. This value means there is a less than a .001 probability of obtaining an *F*-value as high as 137.447 by chance when TREAT has no impact on SALES. In other words, there is less than one chance in a thousand of committing a Type 1 error by rejecting H_o on the basis of the available evidence.

Some additional explanation of the output in Exhibit 19.1 is in order. Perhaps you are wondering why there are three identical rows of numbers (the first three rows) in the computer output. While there was just one independent variable in our illustration, we can examine the impact of two or more independent variables by using ANOVA. The row labeled MAIN EFFECTS refers to results aggregated over *all* independent variables considered in a given situation. Since TREAT was the only independent variable in this case, numbers in the first and second rows of the computer output are the same. (Had there been more than one independent variable, there would have been additional rows of numbers under the row corresponding to TREAT, one row for each additional variable.) The row labeled EXPLAINED refers to the total of individual, or ***main, effects*** of independent variables as well as any effects due to *interactions* between them (interaction effects are described and a computer output containing such effects is presented later). Again, since our example involves only one independent variable, there is no interaction effect. Hence the numbers in the TREAT and EXPLAINED rows are identical. The RESIDUAL row corresponds to variance in the dependent variable not accounted for by the independent variables and their interactions. So the sum of squares and degrees of freedom in the EXPLAINED and RESIDUAL rows must add to the corresponding numbers in the TOTAL row (verify this result in Exhibit 19.1).

Main Effect *The distinct effect of an independent variable on the dependent variable in an ANOVA procedure, regardless of the levels of other independent variables.*

Now that we know TREAT had a significant impact on SALES, which specific treatment group means are significantly different? Visual inspection of the three treatment group means suggests that each may differ significantly from the other two. An extension of the ANOVA technique—called constructing and evaluating *contrasts*—is available to verify the statistical significance of differences between any pair of treatment means. We will not cover the details of this extension.[1] Most computer programs for performing ANOVA offer an option for evaluating specific treatment differences.

The similarity between the expressions for the total sum of squares (SS_T) under ANOVA and regression analysis was pointed out earlier. The appendix to this chapter discusses further the correspondence between the two techniques with the help of dummy-variable regression, which is used when one or more independent variables are categorical.

ADVANCED ANOVA PROCEDURES

Completely Randomized ANOVA *An ANOVA procedure in which study units are randomly assigned to the treatments without explicitly matching the units across the treatments.*

The ANOVA procedure illustrated in the preceding section is a rather basic one called a *one-way,* ***completely randomized ANOVA***. The "one-way" descriptor refers to the fact that only one independent variable—price, with three treatment levels—was involved. Sophisticated extensions of this basic ANOVA procedure are available for analyzing data from the more advanced experimental designs described in Chapter 9. In this section we will briefly describe several such extensions.[2]

---

[1] A good discussion of contrasts is given in William C. Guenther, *Analysis of Variance* (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

Randomized-Block ANOVA

One limitation of the completely randomized experimental design used in the grocery store chain example is that no explicit effort was made to control for the potential effects of extraneous factors (e.g., store size) on unit sales of the vegetable oil. Although random allocation of the sample stores to the three treatments should even out such effects, there is no guarantee that it would be successful in doing so. For instance, stores allocated to the 50¢-off and 75¢-off treatments can, by chance, be larger than those allocated to the regular-price treatment. If so, the observed differences in mean sales across the three groups of stores cannot be unambiguously attributed to the pricing treatments. One way of guarding against such a possibility is to use a randomized-block experimental design.

As we saw in Chapter 9, in a randomized-block design the stores will not be allocated to three treatments on a completely random basis. Rather, they will first be arranged according to size. The three largest stores will then be assigned randomly to the three treatments. Thus one store from the matched threesome of stores is guaranteed to be in each treatment group. Successive threesomes of stores from the arranged list will be assigned to the three treatments in a similar fashion. The result will be eight blocks of stores (each block representing a different store size category), with stores within each block randomly assigned to the three treatments. Blocking in experimentation is analogous to stratification in sampling. Just as stratified random sampling is statistically more efficient than simple random sampling, the randomized-block design is better than the completely randomized design in terms of detecting differences between treatment groups.

Blocking Factor *A relevant characteristic of study units that serves as the basis for forming matched sets of units (called blocks) in certain experimental studies.*

Randomized-Block ANOVA *An ANOVA procedure in which study units are divided into homogeneous blocks on the basis of a blocking factor and the units in each block are randomly assigned to the treatments.*

The format of the raw data set generated through a randomized-block design will be similar to that of Table 19.2, but with one key difference: Each row of data will now correspond to a set of three stores of roughly equal size. Store size in this case is called a ***blocking factor***. ***Randomized-block ANOVA*** achieves its advantage over completely randomized ANOVA by the way it partitions the total variation in the dependent variable (that is, SS_T). Rather than splitting SS_T simply into between (or treatment) and within (or unexplained) sums of squares, it further splits the within sum of squares into a sum of squares due to the blocking factor (SS_b) and a residual or error sum of squares (SS_e). Therefore in a randomized-block ANOVA,

$$SS_T = SS_{TR} + SS_b + SS_e.$$

---

[2] An excellent reference source for learning more about the various ANOVA procedures in Guenther, *Analysis of Variance*.

The only unexplained variance is hence the error variance (SS_e), which in any given situation will invariably be lower than the unexplained variance if a completely randomized design is to be used (that is, SS_W). Also, the *F*-statistic (ratio of treatment mean square to error mean square) for checking the significance of the treatment effects will be higher when a randomized-block ANOVA is used. So the likelihood of rejecting the null hypothesis of no effect due to the treatment variations, when it is false, will be higher. Thus randomized-block ANOVA is *more powerful* in terms of being able to detect differential impacts due to the treatment levels.

The computer output for a randomized-block ANOVA will look similar to the output in Exhibit 19.1, except that it will list separate sums of squares, degrees of freedom, and mean squares for the treatment variable, blocking factor, and residual. It will also list separate *F*-values for the treatment variable and blocking factor so that the statistical significance of the impact of each on the dependent variable can be ascertained.

Latin-Square ANOVA

Latin-Square ANOVA *An ANOVA procedure in which two blocking factors, each having the same number of levels as the number of treatments, are employed simultaneously.*

Latin-square ANOVA is a procedure that permits simultaneous control of the effects of *two* extraneous factors, and it is useful for analyzing data coming from a Latin-square experimental design (described in Chapter 9). To review this design, consider the grocery store chain example. Suppose unit sales of vegetable oil can be affected by not only store size but also store location. The Latin-square design will be appropriate in this situation to evaluate the true effects of the pricing treatments. One major requirement for using the Latin-square design is that the number of levels of each blocking factor and the number of main treatment groups must be identical. Hence in our example we must have exactly three categories of store size (e.g., large, medium, and small) and three categories of store location (e.g., urban, suburban, and rural).

From the three levels of each blocking factor a 3 × 3 matrix of 9 cells, representing all possible combinations of blocking-factor levels, is created. Next, starting with one cell at random, we assign pricing treatments to stores in the various cells such that each treatment appears once, and only once, in each row and column of the matrix. Unit sales are then measured in each cell and analyzed to ascertain the impact of the pricing treatments.

In a Latin-square ANOVA the part of SS_T not attributable to the treatments are split *three* ways: a sum of squares due to the row blocking factor (SS_r), a sum of squares due to the column blocking factor (SS_c), and a residual, or error, sum of squares (SS_e). Therefore

$$SS_T = SS_{TR} + SS_r + SS_c + SS_e.$$

Since in a Latin-square ANOVA the effects of two blocking factors are filtered out of the unexplained variance, the residual (or error) sum of squares (SS_e) is likely to be further reduced and the power to detect differential treatment effects increased, in comparison with the randomized-block procedure. The computer output for a Latin-square ANOVA will be similar to that for a randomized-block ANOVA. It will enable one to perform separate statistical-significance testing of the impact of the row blocking factor, the column blocking factor, and the main-treatment variable.

Factorial ANOVA

Factorial ANOVA An ANOVA procedure in which two or more categorical independent (treatment) variables are involved.

The ***factorial ANOVA*** is used to analyze data from a factorial-design experiment—that is, one intended to investigate the impact of *more than one* categorical independent variable on a metric-scaled dependent variable. A factorial design is thus a multiway design. We will review and illustrate this design by using our grocery chain example. Suppose we want to investigate the impact of an in-store, point-of-purchase display on unit sales of vegetable oil in addition to the pricing factor already described. Also, suppose the presence and the absence of the point-of-purchase display are the two treatment levels for this new factor. When we consider the two factors simultaneously, six *combinations* of treatment levels are possible: presence or absence of the point-of-purchase display with each of the three pricing treatments (regular price, 50¢ off, and 75¢ off). Sample stores can now be allocated to the six treatment levels, and unit sales can be monitored and analyzed to provide insight into the impact of the two factors.

When several factors, each with several levels, are involved in a factorial design, the number of possible combination levels and hence treatment groups can become numerous. In this case the cost of studying a subsample of units under each combination level may be prohibitive. Fortunately, procedures are available for systematically selecting and studying just a few of the levels and still being able to gain insight into the impact of all the factors.[3]

In our example we can, of course, use two separate one-way designs—one to study the impact of price and the other to study the impact of the point-of-purchase display. However, a major limitation of this approach is that it will not indicate possible interaction effects of the two factors on unit sales. An ***interaction effect*** is said to be present when the relative impact of treatment levels of one factor varies across treatment levels of another factor.

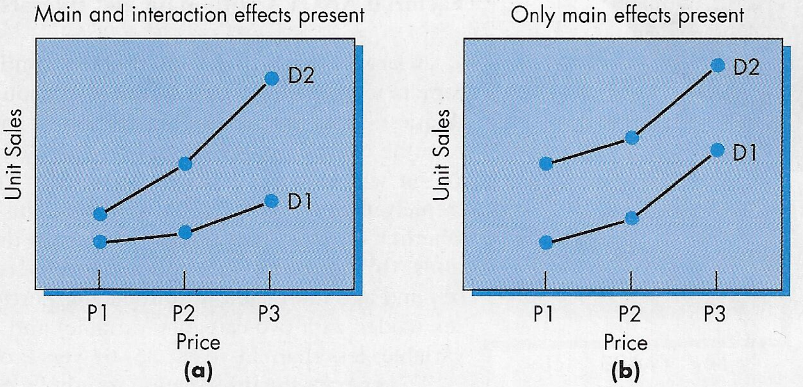
Interaction Effect Effect in an experiment when the relative impact of different levels of one independent variable depends on the levels of another independent variable.

For instance, suppose the increases in unit sales due to the price reductions are much more pronounced when the point-of-purchase display is present than when it is absent. In this case, the two factors (i.e., price and point-of-purchase display) are said to interact with each other in influencing the dependent variable (i.e., unit sales). Exhibit 19.2(a)

---

[3] See, for example, Charles W. Holland and David W. Cravens, "Fractional Factorial Experimental Designs in Marketing Research," *Journal of Marketing Research*, 10 (August 1973), pp. 270–276.

Exhibit 19.2 *Illustrations of Main and Interaction Effects*



Key

P1: Regular price
P2: 50¢ off
P3: 75¢ off
D1: Display absent
D2: Display present

illustrates this interaction effect. In Exhibit 19.2(a) price reductions have a positive impact on unit sales under both the display and no-display treatments. This impact reflects a *main* effect due to the price factor. Similarly, the higher unit sales in the display treatment than in the no-display treatments (under all three price levels) reflect a *main* effect due to the point-of-purchase display factor. The *steeper* increases in unit sales (in response to price reductions) under the display treatment than under the no-display treatment, and the resulting nonparallel sales-trend lines, reflect the *interaction* effect due to the joint influence of price and point-of-purchase display.

Exhibit 19.2(b) depicts a situation wherein there are main effects but no interaction effect. Price reductions result in higher sales under both the display and no-display treatments, and the display treatment results in higher sales than the no-display treatment under all three price levels. However, the increases in sales due to price reductions are the same for the display and no-display treatments (i.e., the slopes of corresponding sections of the two sales-trend lines are parallel). Moreover, the increases in sales due to having a display are the same for the three price treatments (i.e., gap between the two sales-trend lines is the same at all three price levels). Therefore, there is no interaction effect.

In factorial ANOVA the sum of the squares across all treatment combinations (six such combinations in our example), denoted as SS_{TR}, is split into sums of squares attributable to each treatment variable separately and to interactions between treatment variables. If we label the treatment variables in a two-way ANOVA as A and B, we can represent

Research in Use 19.1 *Bank Customer Perceptions Study*

Factorial ANOVA with Main and Interaction Effects

A large bank in the United States conducted a mail survey of its customers to assess their perceptions of various aspects of the bank's services. A question in the concluding section of the survey asked respondents to rate the overall quality of service provided by the bank on a scale of 1 to 10 on which 1 was labeled "Extremely Poor" and 10 was labeled "Extremely Good." The bank's management was interested in ascertaining whether (a) male and female customers differed in their overall perceptions; (b) customers' perceptions differed according to their ages; and (c) sex and age interacted in influencing perceptions. Data were available on sex (coded as a two-category variable) and age (coded as a three-category variable: less than 35 years; 35–64 years; over 64 years).

To generate the information sought by management a two-way factorial ANOVA was conducted. In this ANOVA sex and age were the two categorical independent variables and overall quality rating was the metric-scaled dependent variable. Of 372 customers who responded to the survey, 335 had data on all three variables. Therefore the total sample size for the analysis was 335. The computer output from using the SPSS ANOVA program is reproduced next.

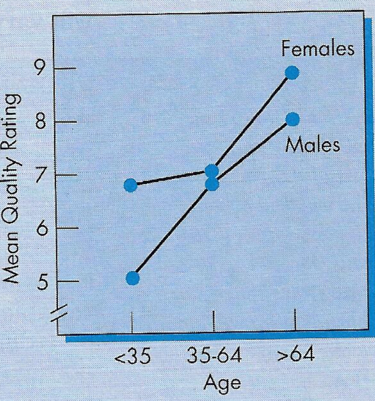
```
                              ANALYSIS OF VARIANCE

QUALITY BY
AGE/SEX

SOURCE OF          SUM OF                 MEAN                      SIGNIF
VARIATION          SQUARES       DF       SQUARE          F         OF F

MAIN EFFECTS        302.434       3       100.811      21.659       0.000
   AGE              283.893       2       141.946      30.497       0.000
   SEX               42.603       1        42.603       9.153       0.003
2-WAY
INTERACTIONS         29.951       2        14.976       3.218       0.041
   AGE     SEX       29.951       2        14.976       3.218       0.041
EXPLAINED           332.385       5        66.477      14.283       0.000
RESIDUAL           1531.305     329         4.654
TOTAL              1863.690     334         5.580
   372 CASES WERE PROCESSED.
     37 CASES (9.9 PCT) WERE MISSING.
```

As shown by the computer output, the main effects of both SEX and AGE as well as their interaction effect on QUALITY are highly significant. Notice that the actual significance levels (shown in the last column of the output) for the effects of AGE, SEX, and AGE-SEX interaction are all below the customary significance level of .05. The mean quality ratings (generated by the SPSS ANOVA program) for different subgroups of respondents are summarized in the table (numbers in parentheses are sample sizes for the subgroups). The means for the six SEX-AGE combinations are also portrayed graphically in the accompanying figure.

(Continued)

Research in Use 19.1 *Continued*

AGE	SEX Male	 Female	 Total
<35	5.13 (40)	6.79 (48)	6.03 (88)
35-64	6.85 (73)	7.05 (81)	6.95 (154)
>64	8.14 (58)	8.83 (35)	8.40 (93)
Total	6.88 (171)	7.35 (164)	7.11 (335)



The figure shows a main effect for sex (females have more favorable perceptions than males), a main effect for age (the older the customers the more favorable are their perceptions), and a sex-age interaction effect (notice the *nonparallel* quality-perception-trend lines for males and females). The interaction effect here shows that the difference in quality perceptions between males and females varies by age—the difference is most pronounced for the under-35 age group and least pronounced for the 35–64 age group.

the partitioning of the total sum of squares as follows:

$$\begin{aligned} SS_T &= SS_{TR} + SS_e \\ &= SS_A + SS_B + SS_{AB} + SS_e, \end{aligned}$$

where SS_A and SS_B correspond to the main effects of the two treatments, SS_{AB} corresponds to the interaction effect between them, and SS_e is the error sum of squares.

The computer output for a factorial ANOVA will indicate the value of each SS-component in the expression for SS_T along with its associated degrees of freedom and mean square value. It will also provide separate *F*-values for ascertaining the significance of each main effect and the interaction effect. Research in Use 19.1 describes and presents results from a study in which main and interaction effects were investigated by using the SPSS ANOVA program.

A factorial experimental design is more efficient than several one-way designs in that it costs about the same and yet provides richer information. This design is also quite versatile in terms of allocating units to treatments and incorporating blocking factors. Units can be allocated by using a completely randomized procedure (as in the very first design we looked at), a randomized-block procedure, or a Latin-square procedure; partitioning of the total sum of squares and the resulting ANOVA table will vary accordingly.

Analysis of Covariance (ANCOVA)

Analysis of Covariance (ANCOVA) *An advanced analysis of variance procedure in which the effects of one or more metric-scaled extraneous variables (called covariates) are removed from the dependent-variable data before one conducts the ANOVA.*

Analysis of covariance (ANCOVA) is simply a variation of the ANOVA procedures and is intended to remove the effects of extraneous variables on the dependent variable. ANCOVA is especially suitable for situations where one cannot explicitly control an extraneous variable, as is possible in, say, a randomized-block design. One key requirement of ANCOVA is that metric data must be available for sample units on each extraneous variable to be included in the analysis.

To illustrate what is involved in ANCOVA, let us once again look at the one-way, completely randomized experimental design for investigating the effect of price on unit sales. As we saw earlier, one extraneous variable that could influence unit sales in addition to price is store size. Suppose metric data are available on store size (e.g., square feet of area), but for some reason we cannot use a randomized-block design to block out the effect of this extraneous factor. With metric data on unit sales and store size for all the sample stores, we can treat the former as a dependent variable and the latter as an independent variable and perform a regression analysis. Using the resulting regression equation, we can predict unit sales for each sample store and also compute a *residual*, representing the difference between the actual and predicted unit sales. We can thus generate a set of residual values for stores in each treatment group.

Covariate *An extraneous variable whose effects on the dependent variable are removed in conducting an analysis of covariance.*

What do the residual values represent? They are simply the portion of unit sales not accounted for by store size. Stated differently, the residuals are data on the dependent variable that are no longer contaminated by the effects of store size. Therefore by analyzing these residuals (rather than the raw unit sales) through ANOVA, we should get a clearer picture of the pricing-treatment effects.

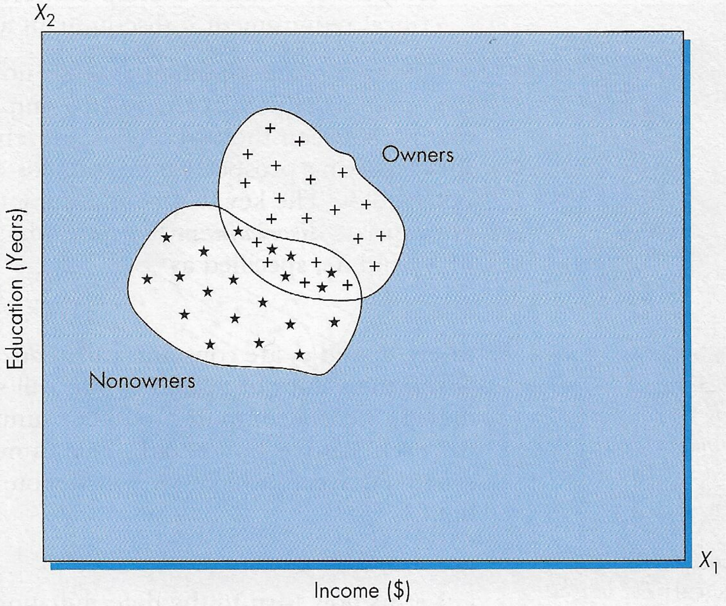
The modified ANOVA procedure illustrates the essence of ANCOVA. Stated simply, ANCOVA is ANOVA performed on dependent-variable residuals generated by first conducting a regression analysis wherein the extraneous variables are the independent variables. The extraneous factors included in an ANCOVA procedure are appropriately called ***covariates***.

DISCRIMINANT ANALYSIS

Discriminant Analysis *A technique for generating a linear combination of independent variables (called a discriminant function) in such a way that values on it (called discriminant scores) are similar for sample units within prespecified subgroups and are maximally separated across the subgroups. The prespecified subgroup categories make up the categorical dependent variable in discriminant analysis.*

Discriminant analysis deals with identifying the distinguishing features of prespecified subgroups of units that are formed on the basis of some dependent variable (e.g., heavy, moderate, and light users of a product; homeowners and renters; viewers and nonviewers of a television program). Thus the dependent variable in discriminant analysis is categor-

Exhibit 19.3 *Scatter Plot of Income and Education Data for Personal-Computer Owners and Nonowners*



ical, and there are as many prespecified subgroups as there are categories. The independent variables are typically metric-scaled. The purpose of discriminant analysis includes classification of new units into one of the subgroups, given the new units' values on the independent variables.

To illustrate discriminant analysis, consider the following example.

Example. A computer manufacturer wants to see whether household income (X_1) and the number of years of formal education of the household head (X_2) are useful in clearly distinguishing households owning personal computers from those not owning personal computers. If X_1 and X_2 seem to be crucial determinants of personal-computer ownership, the firm also wishes to be able to determine whether a prospective customer household is likely to buy a personal computer, given the household's values on X_1 and X_2. The firm has gathered data on X_1 and X_2 for two random samples of households, an owner sample and a nonowner sample. These data are plotted on the two-dimensional graph shown in Exhibit 19.3.

Two enclosed areas are shown in Exhibit 19.3, one containing primarily households with personal computers and the other containing primarily households without personal computers. Although the two areas overlap, the extent of overlap does not seem to be substantial. Intuitively, this exhibit implies that income and education are good discriminators between households with and those without personal computers. Minimal overlap between groups, as in Exhibit 19.3, is a critical requirement if discriminant analysis is to be fruitful.

The task of discriminant analysis now is to utilize the two groups of data points to quantify the relative importance of X_1 and X_2 in discriminating between the two sets of households, and to develop a criterion for classifying prospective households as potential owner or nonowner households. The key to accomplishing this twofold task is to develop an appropriate ***discriminant function***, which is simply a linear combination of X_1 and X_2, specified as

$$v_1X_1 + v_2X_2,$$

where v_1 and v_2 are constants called ***discriminant weights***, or ***coefficients***. Given the values of v_1 and v_2 (we will see shortly how these are determined), we can compute a number summarizing the linear combination for each sample household. This number is called a household's ***discriminant score***, which we will denote as its Y-score. For any household h,

$$Y_h = v_1X_{1h} + v_2X_{2h}.$$

Discriminant Function *A mathematical expression representing the linear combination of independent variables employed in a discriminant analysis.*

Discriminant Weight (Coefficient) *The weight (coefficient) corresponding to an independent variable in the discriminant function.*

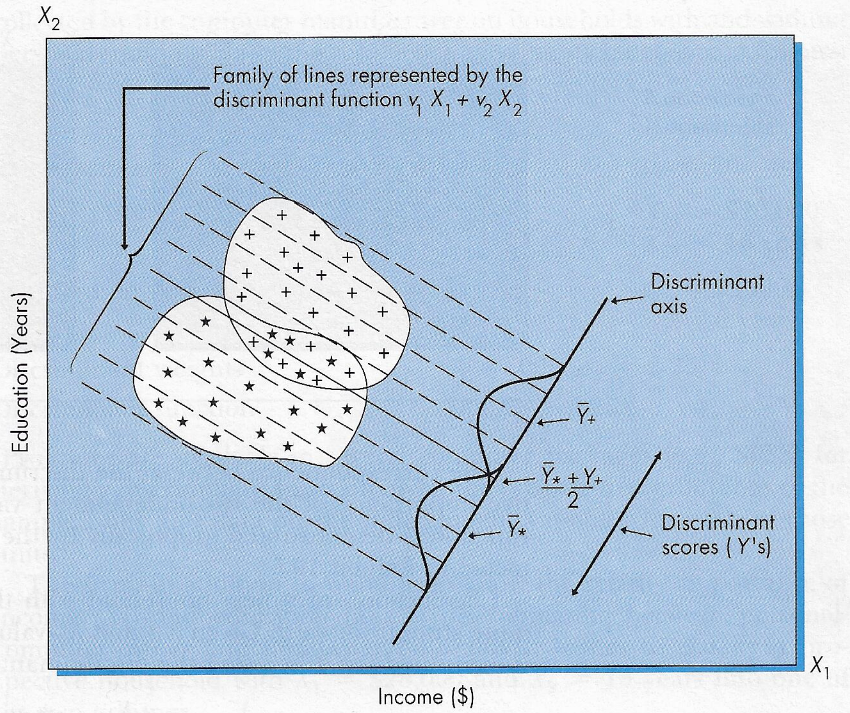
Discriminant Score *The discriminant-function value obtained for a unit by inserting the unit's independent-variable values into the discriminant function.*

Let us now turn to the determination of v_1 and v_2. Theoretically, an infinite number of combinations of v_1- and v_2-values are possible. Furthermore, depending on these values, the Y-scores for the sample households will vary. What discriminant analysis does is to select v_1 and v_2 in such a way that the variation in the Y-scores *between* the two groups of households is made as large as possible relative to variation in the Y-scores *within* them. In other words, v_1 and v_2 are chosen to maximize the ratio of between to within sums of squares corresponding to Y-scores. The discriminant function is thus the linear combination of the independent variables that offers the best separation (least overlap) of Y-scores between the prespecified groups.

A graphical interpretation of a discriminant function of the form $v_1X_1 + v_2X_2$ is as follows: It is a family of parallel straight lines in a two-dimensional graph with X_1 and X_2 as axes. Each straight line in the family corresponds to a specific Y-score. Therefore the Y-scores of data points on the same straight line will be identical, but Y-scores of data points on different straight lines will vary. In fact, we can measure the Y-scores along a line perpendicular to the family of lines representing the discriminant function. This perpendicular line is called the ***discriminant axis*** and represents the dimension along which maximum separation between the two groups occurs. Stated differently, projections on the discriminant axis of the data points corresponding to the two groups will be clustered together within each group and spread apart between groups. The graphical interpretation of the discriminant function and discriminant scores is illustrated in Exhibit 19.4.

Discriminant Axis *A line that is perpendicular to the family of lines representing the discriminant function.*

Exhibit 19.4 *Discriminant Function and Discriminant Scores*


X_2
Family of lines represented by the discriminant function $v_1 X_1 + v_2 X_2$
Discriminant axis
$\bar{Y}_+$
$\frac{\bar{Y}_* + \bar{Y}_+}{2}$
$\bar{Y}_*$
Discriminant scores (Y's)
Education (Years)
Income ($)
X_1

Using the Discriminant Function

The weights v_1 and v_2 can be interpreted as signifying the relative importance of X_1 and X_2 in being able to discriminate between the two groups, provided both independent variables are measured in the same units. When the measurement units are different (as in our example), one must obtain standardized weights, say v_1^* and v_2^*, as follows:

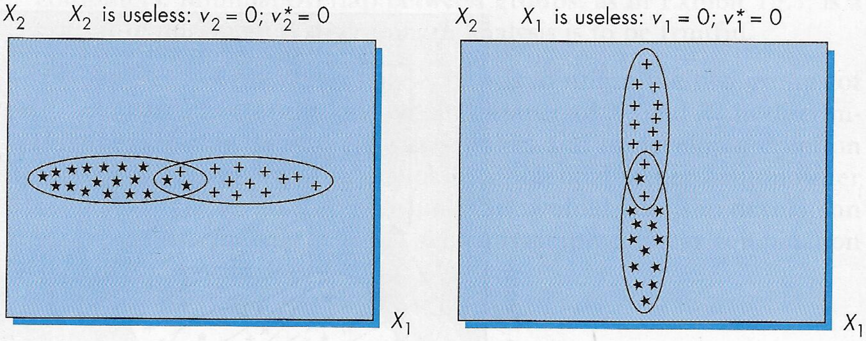
$$v_1^* = v_1 s_1, \qquad v_2^* = v_2 s_2,$$

where s_1 and s_2 are *pooled standard deviations* for X_1 and X_2, respectively. The pooled standard deviation s_k for any variable X_k can be computed by using the following formula in which the subscripts 1 and 2 refer to the two samples:

$$s_k = \sqrt{\frac{(n_1 - 1)s_{k1}^2 + (n_2 - 1)s_{k2}^2}{n_1 + n_2 - 2}}.$$

Use of this formula requires the assumption of equal X_k variances in the populations implied by the two samples.

Exhibit 19.5 *Relative Importance of Independent Variables in a Discriminant Function*


X_2
X_2 is useless: $v_2 = 0$; $v_2^* = 0$
X_1
X_2
X_1 is useless: $v_1 = 0$; $v_1^* = 0$
X_1

The correspondence between the discriminant weights and the relative importance of the two independent variables in the discriminant function is demonstrated graphically by the two extreme situations depicted in Exhibit 19.5.

Classification of a new household with the discriminant function is quite straightforward. Given X_1- and X_2-values for a new unit, a Y-score is computed for it by using the discriminant function:

$$Y_{\text{new}} = v_1 X_{1,\text{new}} + v_2 X_{2,\text{new}}.$$

Then Y_{new} is compared with the mean discriminant scores, $\overline{Y}_*$ and $\overline{Y}_+$, for the two groups, which are computed as follows (the subscript + refers to personal-computer-owner households and the subscript * refers to nonowner households):

$$\overline{Y}_* = v_1 \overline{X}_{1*} + v_2 \overline{X}_{2*},$$
$$\overline{Y}_+ = v_1 \overline{X}_{1+} + v_2 \overline{X}_{2+}.$$

The new household is assigned to the owner group if Y_{new} is closer to $\overline{Y}_+$ than to $\overline{Y}_*$; if not, it is assigned to the nonowner group. Another way of stating this assignment criterion is as follows: Compare Y_{new} with a *critical discriminant score*, Y_{cri}, defined as

$$Y_{\text{cri}} = \frac{\overline{Y}_* + \overline{Y}_+}{2}.$$

If $Y_{\text{new}} > Y_{\text{cri}}$, assign the household to the owner group; it not, assign it to the nonowner group. This allocation rule implicitly assumes that the cost of misallocation is the same for both groups. That is, erroneously classifying a potential owner as a nonowner will cost the same as classifying a potential nonowner as a owner. When the costs of misallocation are different, a more sophisticated classification criterion (not discussed here) is needed.[4] The solid line in the family of lines in Exhibit 19.4 is the critical discriminant line.

To illustrate the use of a discriminant function, suppose the data collected by the computer manufacturer on households with and without personal computers and the results of analyzing such data are as follows:

	PC-Owner Households	**Nonowner Households**
Sample sizes	$n_+ = 20$	$n_* = 20$
Sample means	$\bar{X}_{1+} = \$30{,}000$ $\bar{X}_{2+} = 18$ years	$\bar{X}_{1*} = \$20{,}000$ $\bar{X}_{2*} = 16$ years
Standard deviations	$s_{1+} = \$6000$ $s_{2+} = 4$ years	$s_{1*} = \$5000$ $s_{2*} = 6$ years
Discriminant weights	$v_1 = .009;\ v_2 = 3.25$	
Discriminant function	$.009X_1 + 3.25X_2$	

[Programs are available in several computer packages (e.g., SPSS) for deriving a discriminant function given the prespecified groupings of the sample units and data on the independent variables for each of those units.]

The firm, in addition to wanting to know the relative importance of income (X_1) and education (X_2) in discriminating between personal-computer-owner and nonowner households, wishes to classify a prospective household with $X_1 = \$28{,}000$ and $X_2 = 15$ years into one of the two groups.

Since X_1 and X_2 are measured in different units, the discriminant weights v_1 and v_2 must be standardized through multiplication by the corresponding pooled standard deviations s_1 and s_2:

$$s_1 = \sqrt{\frac{(n_+ - 1)s_{1+}^2 + (n_* - 1)s_{1*}^2}{n_+ + n_* - 2}}$$

$$= \sqrt{\frac{(19)(6000)^2 + 19(5000)^2}{38}} = \$5522.68,$$

$$s_2 = \sqrt{\frac{(n_+ - 1)s_{2+}^2 + (n_* - 1)s_{2*}^2}{n_+ + n_* - 2}}$$

$$= \sqrt{\frac{(19)(4)^2 + 19(6)^2}{38}} = 5.10 \text{ years},$$

$$v_1^* = s_1 v_1 = (5522.68)(.009) = 49.70,$$

$$v_2^* = s_2 v_2 = (5.10)(3.25) = 16.52.$$

---

[4] For details about this classification criterion and other aspects of discriminant analysis, see Donald F. Morrison, "Discriminant Analysis," in *Handbook of Marketing Research*, ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 2.442–2.457.

Comparing v_1^* with v_2^*, we can say that household income is about three times as important as education of the household head in distinguishing between personal-computer-owner and nonowner households.

To classify the prospective household, we need to determine its Y-score and compare it with Y_{cri}:

$$\begin{aligned}
\bar{Y}_{\text{new}} &= v_1 X_{1,\text{new}} + v_2 X_{2,\text{new}} \\
&= (.009)(28{,}000) + (3.25)(15) \\
&= 252 + 48.75 = 300.75, \\
\bar{Y}_{+} &= (.009)(30{,}000) + (3.25)(18) \\
&= 270 + 58.5 = 328.5, \\
\bar{Y}_{*} &= (.009)(20{,}000) + (3.25)(16) \\
&= 180 + 52 = 232, \\
Y_{\text{cri}} &= \frac{\bar{Y}_{+} + \bar{Y}_{*}}{2} = \frac{328.5 + 232}{2} = 280.25.
\end{aligned}$$

Since $Y_{\text{new}} > Y_{\text{cri}}$, we should classify the prospective household as a potential owner of a personal computer.

Evaluating a Discriminant Function

The preceding discussion of discriminant analysis has not considered the issue of how trustworthy the resulting discriminant function is. However, tests are available for checking the statistical significance of a discriminant function as well as its discriminant weights. These tests are similar to the ones we discussed for regression analysis.[5] Since both regression analysis and discriminant analysis involve forming linear combinations of independent variables, the potential limitations of regression analysis and the precautions to be taken in interpreting its results are applicable to discriminant analysis as well.

Confusion Matrix *A table set up to evaluate the predictive accuracy of a discriminant function. It shows the degree of correspondence, or lack thereof, between the actual groupings of the sample units and the predicted groupings obtained by classifying the same units through the discriminant function.*

For a discriminant function that is statistically significant, an intuitive approach for evaluating its practical usefulness is to construct and examine what is usually called a confusion matrix. A ***confusion matrix*** indicates the degree of correspondence, or lack thereof, between the actual groupings of the sample units and the predicted groupings obtained by classifying the same units through the discriminant function. The confusion matrix corresponding to our illustration (Exhibit 19.4) is shown in Table 19.3.

Numbers on the diagonal going down from left to right in the confusion matrix refer to sample units correctly classified. A summary indicator of a discriminant function's predictive ability is the following ratio, called the ***hit rate:***

$$\frac{\text{total number of correctly classified units}}{\text{total sample size}}.$$

Hit Rate *The porportion of units correctly classified by a discriminant function.*

---

[5] For further details, see Morrison, "Discriminant Analysis."

Table 19.3 *Confusion Matrix Corresponding to Exhibit 19.4*

	Predicted groupings	
Actual Groupings	**Households with Personal Computers**	**Households without Personal Computers**
Households with personal computers	17	3
Households without personal computers	4	16

The hit rate for the confusion matrix in Table 19.3 is

$$\frac{17 + 16}{20 + 20} = \frac{33}{40} = .825.$$

Thus our illustrative discriminant function is able to correctly classify 82.5% of all sample households.

Note that the hit rate calculated here is based on the same set of sample units used to construct the discriminant function. Hence it is likely to be somewhat biased in favor of the function. A more stringent way to measure a discriminant function's predictive ability is to construct the function by using part (say, one-half or two-thirds) of the sample and to compute the hit rate on the basis of predictions for the units in the remainder of the sample. Being able to use this approach, of course, depends on how large the sample size is.

How good is the hit rate (prediction accuracy) of 82.5%? The answer depends on how accurately one can classify the units by chance (i.e., without using the discriminant function). The extent to which correct classifications can be made purely by chance is a function of the relative sizes of the two actual groups. In our example both groups are of the same size (i.e., 20), implying that the a priori probability of any sample unit being a personal-computer owner is .5. Also, our objective is to classify the units into owners and nonowners. So suppose we toss a fair coin for each unit and classify the unit as a personal-computer owner if we observe heads (nonowner, if we observe tails). We will be correct 50% of the time, and therefore our prediction accuracy by chance will be 50%. Therefore the hit rate of 82.5% for our discriminant function appears to be quite good.

What if the two groups are not of the same size? For instance, suppose the nonowner group is three times as large as the owner group (i.e., $n_+ = 10$ and $n_* = 30$)? Now the a priori probability of a sample unit being a personal-computer owner is .25. Therefore if we want to classify each unit as either an owner or a nonowner without the aid of a discriminant function, we can use the following procedure: For each unit, draw a random number between 1 and 100. If the number is between 1 and 25, classify the unit as an owner; if not, classify the unit as a nonowner. The chance prediction accuracy in this case can be shown to be

$$P^2 + (1 - P)^2,$$

where P is the proportion of units in one group and $(1 - P)$ is the proportion of units in the other. When $P = .25$, the chance prediction accuracy is $(.25)^2 + (.75)^2 = .625$, or 62.5%. A hit rate of 82.5% compared with a chance prediction accuracy of 62.5% is still not bad, although it is worse than it is when the two groups are of equal size—that is, $P = .5$ (verify that the chance prediction accuracy is 50% when $P = .5$).

Proportional Chance Criterion *The proportion of correct classifications that would result if each unit in a sample is probabilistically classified into one of two prespecified groups solely on the basis of the relative sizes of the two groups.*

Maximum Chance Criterion *The proportion of correct classifications that would result if all sample units are classified into the larger of two prespecified groups.*

In general, the more dissimilar the sizes of the two original groups, the lower is the ability of a discriminant function to outperform a chance classification scheme that relies merely on the relative group sizes. Such a chance classification scheme implies the use of what is usually called the ***proportional chance criterion***. There is another chance classification scheme that uses the so-called ***maximum chance criterion***, in which *all* the units are classified into the larger of the two a priori groups. Thus when a sample of 40 units has 10 owners and 30 nonowners, classifying all 40 units as nonowners will result in a maximum prediction accuracy of 75%. The maximum chance criterion is unrealistic, however, since putting all units into just one group defeats the very purpose of identifying distinct group memberships.

Final Remarks

Our treatment of discriminant analysis centered on situations where the total sample is divided into two groups on the basis of some prespecified criterion. However, discriminant analysis is not limited to just two-group situations. It is capable of generating discriminant functions for classifying units into more than two prespecified groups. The rationale underlying multiple-group discriminant analysis is similar to that underlying the two-group case, although details of the technique are much more complex.

Discriminant functions are not restricted to just two independent variables either. The sole reason for including only two independent variables in our discussion was to keep it simple and to facilitate graphical interpretation of what the technique does. Discriminant functions can and often do contain more than two independent variables. The general form for a discriminant function with k variables is

$$v_1X_1 + v_2X_2 + \cdots + v_kX_k.$$

Discriminant analysis has been used in a wide variety of settings, although it is not as popular as regression analysis or analysis of variance.[6]

---

[6] See Barnett A. Greenberg, Jac L. Goldstucker, and Danny N. Bellenger, "What Techniques Are Used by Marketing Researchers in Business? *Journal of Marketing*, 41 (April 1977), pp. 62–68; A. Parasuraman, "A Study of Techniques Used and Clients Served by Marketing Research Firms," *European Research*, 10 (October 1982), pp. 177–185.

The potential for applying discriminant analysis is great, however, given marketers' ever-present need for defining customer segments, identifying critical characteristics capable of distinguishing among them, and classifying prospective customers into appropriate segments.[7] Rumor has it that even the Internal Revenue Service has developed a sophisticated discriminant function to help classify tax returns into those worthy of an IRS audit and those that are not.

FACTOR ANALYSIS

Factor Analysis *A technique that analyzes data on a relatively large set of variables and produces a smaller set of factors, which are linear combinations of the original variables, so that the set of factors captures as much information as possible from the original data set.*

Factor analysis is essentially a data and variable reduction technique that attempts to partition a given set of variables into groups of maximally correlated variables. For metric-scaled data on a large number of variables, factor analysis generates a smaller number of variables, called *factors,* which captures as much information as possible from the original data set. The factors are formed by taking advantage of interrelationships among the original variables. Factor analysis typically begins by examining a matrix of pairwise correlations among the original variables and explores ways of combining them into factors such that each factor primarily represents a group of maximally correlated variables. The mechanics of analyzing correlations to derive the factors will not be covered here.[8] Rather, our primary focus will be on providing an intuitive understanding of what factor analysis does and how its results can be interpreted and used.

Intuitive Explanation

To examine the procedure of factor analysis, let us first consider this hypothetical situation.

> **Example.** Star Brands, Inc. (SBI), a manufacturer of a variety of home appliances and electronic products, recently conducted a survey of customers owning Star products. The purpose of the survey was to find out how customers felt about SBI in general and about specific Star products in particular. The survey involved numerous evaluative statements that respondents answered by using a seven-point scale on which the higher the number, the more favorable was the evaluation.

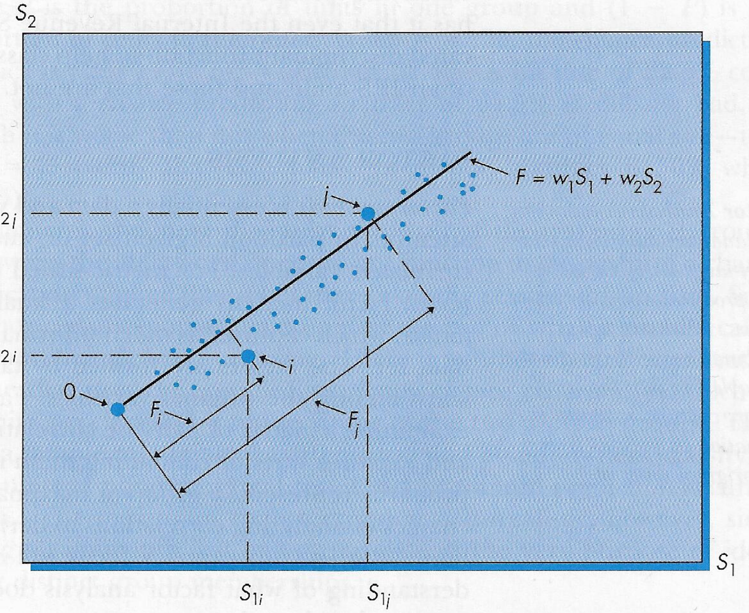
A variety of rating scales were discussed in Chapter 11. Recall that data obtained through rating scales are typically assumed to have interval-scale properties. Factor analysis requires data that are at least interval-scaled.

---

[7] One marketing research firm specializes in using discriminant analysis to advise clients on what characteristics distinguish likers of their brands from dislikers, and how to reposition the brands to convert dislikers into likers. See Lorin Zissman and John Morton, "Brand Maximization Analysis Clarifies Product Repositioning Opportunities," *Marketing News,* January 22, 1982, Section 1, pp. 4–5.

[8] The following texts provide good discussions of the technical details concerning factor analysis: Jum C. Nunnally, *Psychometric Theory* (New York: McGraw-Hill, 1978); Harry H. Harman, *Modern Factor Analysis* (Chicago: University of Chicago Press, 1967).

Exhibit 19.6 *Situation Where Factor Analysis Will Be Beneficial: S_1 and S_2 Highly Correlated*



Although factor analysis typically involves analyzing data on a large number of statements (variables) simultaneously, let us just consider the following two statements from the SBI survey in order to illustrate graphically the rationale underlying the technique:

1. I have been satisfied with the Star products I have purchased.
2. When I have to purchase a home appliance in the future, it will likely be a Star product.

Customer ratings on each statement can be considered as data on a single variable. Let S_1 and S_2 denote the variables implicit in the two statements. A plot of the survey data on S_1 and S_2 is shown in Exhibit 19.6.

Clearly, S_1 and S_2 are highly correlated. In other words, ratings on the two statements are largely redundant since they apparently represent very similar customer sentiments. Therefore we should be able to combine them into ratings along just one dimension without any significant loss of information embedded in the raw data set. This combination is exactly what factor analysis of the data on S_1 and S_2 will accomplish. The straight line labeled F in Exhibit 19.6 is a factor and represents the dimension along which measurements can be made for all sample units

Factor Score *A term used in factor analysis; it represents the value for a sample unit on a given factor and is obtained by substituting the sample unit's values on the original variables into the linear combination representing the factor.*

from an arbitrary origin such as point 0 shown on the line. These measurements are called ***factor scores***. Two such factor scores, F_i and F_j, are shown in the exhibit for respondents i and j.

What does a comparison of the factor scores for respondents i and j with their ratings on S_1 and S_2 reveal? Clearly, the higher a respondent's rating on S_1 *or* S_2, the higher the respondent's factor score is in an almost proportionate fashion. Exhibit 19.6 further suggests that the close correspondence between the two original sets of ratings and the factor scores can be generalized to the entire sample. In factor analysis terminology S_1 and S_2 are said to have high factor loadings on F.

Definition. A ***factor loading*** is the Pearson correlation coefficient between an original variable and a factor.

In short, having data on just F for the sample respondents is virtually as good as having separate ratings on S_1 and S_2. What Exhibit 19.6 demonstrates graphically for just two variables is the foundation underlying factor analysis and can be generalized to more than two variables. Given data on a large number of variables, *at least some of which are highly correlated with one another,* factor analysis can provide a more parsimonious set of factors with little loss of information. The emphasis on correlations among the original variables in the preceding sentence is noteworthy: Factor analysis will not do much good if the original variables are poorly correlated.

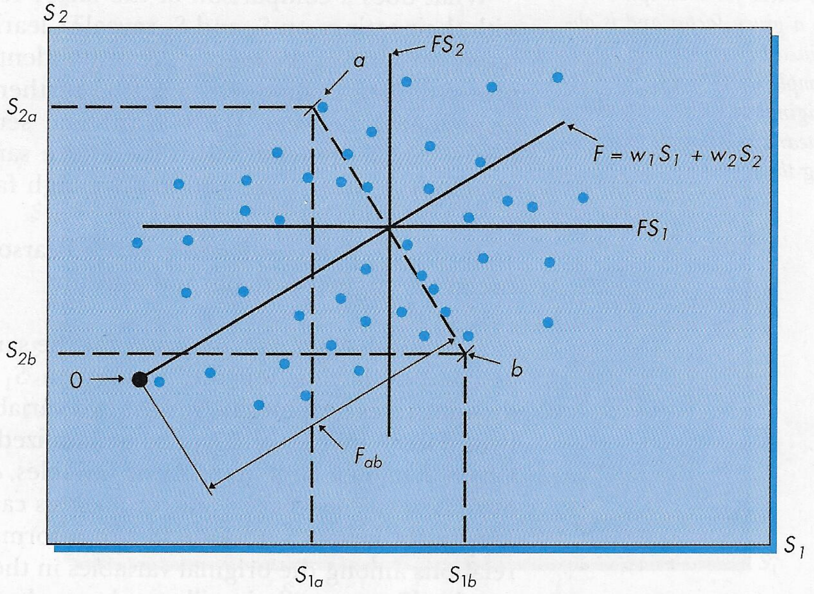
This limitation of factor analysis is demonstrated in Exhibit 19.7, in which S_1 and S_2 are poorly correlated. Poor association between factor scores and S_1 and S_2 ratings is illustrated by the fact that all data points along line ab (with S_1 ratings ranging from S_{1a} to S_{1b} and S_2 ratings ranging from S_{2a} to S_{2b}) will have the same factor score F_{ab}.

The factor scores along dimension F in Exhibit 19.7 do not correlate strongly with the data on either S_1 or S_2. The factor F thus does a poor job of capturing the essence of the original data set. We can make F parallel to the S_1 axis, as shown by line FS_1, so that the factor scores will perfectly correspond with the S_1 ratings. Unfortunately, factor scores along FS_1 will reflect nothing about the ratings on S_2. We would have to construct another perpendicular factor, such as FS_2, to capture the S_2 ratings. But by doing so, we would defeat the purpose of factor analysis, namely, data and variable reduction. Obviously, using the original two variables themselves is sensible if two factors are needed to fully capture the gist of the raw data. Thus factor analysis is not always guaranteed to accomplish its intended purpose.

Returning to Exhibit 19.6, what interpretation can we place on factor F? Mathematically, it is a linear combination that accounts for almost all the variation in S_1 and S_2:

$$F = w_1S_1 + w_2S_2,$$

Exhibit 19.7 *Situation Where Factor Analysis Will Not Be Beneficial: S_1 and S_2 Poorly Correlated*



where w_1 and w_2 are weights determined by the factor analysis procedure. But does it have any *substantive* meaning? In other words, since F is a new, combined variable, what *name* can we give it? The issue of naming factors is subjective and hence somewhat controversial. Invariably, users of factor analysis simply choose a label on the basis of any common thread running through the original variables that have high loadings on it. So naming a factor can be difficult when variables with high loadings on it have little in common, which is not an unusual occurrence in factor analysis applications and thus is a potential limitation of the technique. In our illustration S_1 represents satisfaction with Star products and S_2 represents likelihood of future purchases of Star products. Hence we can perhaps consider F as *trust* in Star products. Someone else looking at Exhibit 19.6 may come up with a label other than trust, however.

Factor Analysis Output and Its Interpretation

Canned computer programs are available to perform factor analysis. These programs typically conduct the analysis after standardizing the variables (i.e., converting each to a mean of zero and variance of one). The primary output of such programs is a ***factor-loading matrix***. To

Factor-Loading Matrix
A matrix containing the pairwise correlations between a set of original variables and the set of factors derived through factor analysis.

Table 19.4 *Factor-Loading Matrix Based on Data from Study of Star Customers*

	Factors		
Factor Loadings	F_1	F_2	**Achieved Communalities**
X_1: I did not mind paying the high price for my Star VCR	0.89	0.15	.815
X_2: I am pleased with the variety of things that a Star VCR can do	0.16	0.86	.766
X_3: I hardly ever worry about anything going wrong with my Star VCR	0.18	0.94	.916
X_4: My friends are very impressed with the Star VCR	0.96	0.06	.926
X_5: The Star VCR has the latest technology built into it	0.09	0.88	.782
X_6: No other brand of VCR even comes close to matching the Star VCR	0.92	0.17	.875
Eigenvalues: Standardized variance explained by each factor	2.626	2.454	
Proportion of the total variance explained by each factor	0.438	0.409	

illustrate this matrix, let us assume that a factor analysis is conducted on ratings on a 7-point scale given by owners of Star videocassette recorders (VCRs) on six statements included in the SBI survey. These statements (designated as variables X_1 through X_6) and their factor loading matrix are shown in Table 19.4.

Table 19.4 implies that two factors (F_1 and F_2) have been constructed by analyzing data on variables X_1 through X_6. In general, as many factors as there are variables can be derived from an original data set. Computer programs allow factor analysis users to specify the number of factors to be extracted. Of course, the smaller the number of factors that can adequately reflect the information in the original data set the better.

One important feature of factor analysis that is crucial in interpreting its results is that the factors themselves are *independent*. In other words, the technique constructs factors in such a way that the correlation between any pair of extracted factors will be zero. Thus there is no overlap or redundancy of information among the factors.

The numbers in Table 19.4 under the column heading "Factors," through the sixth row, are factor loadings. Thus, for instance, the correlation coefficient between X_1 and F_1 is .89; that between X_1 and F_2 is .15. Recall from Chapter 18 that the squared correlation coefficient between two variables is the proportion of variance in one accounted for by the other. Therefore $(.89)^2$ is the variance in X_1 accounted for by F_1, and $(.15)^2$ is the variance in X_1 accounted for by F_2. Since there is no duplication of information between F_1 and F_2, the variance in X_1 extracted by both factors considered together is $(.89)^2 + (.15)^2 = .815$. This value is called the achieved communality for X_1. The achieved communalities for all six variables are listed in the last column of Table 19.4.

Definition. The ***achieved communality*** for any original variable represents the proportion of variance in it accounted for by all the extracted factors.

Achieved communalities for the original variables are given by summing the squared factor loadings corresponding to each variable. They shed light on the extent to which a reduced set of factors is able to reflect the data on each original variable. The two factors in our example account for over 75% of the variance in each of the six variables, suggesting thereby that the factor analysis has been quite effective.

Numbers in the row labeled "Eigenvalues" in Table 19.4 summarize the amount of information captured by each factor.

Definition. The ***eigenvalue*** of any factor is the total standardized variance accounted for by the factor.

Eigenvalues of the extracted factors are given by summing the squared factor loadings corresponding to each factor. The total standardized variance in the original data set is simply the number of variables, since each variable has a variance of one. The total standardized variance in our example is 6. Therefore, the proportion of variance extracted by F_1 is its eigenvalue divided by 6, or $2.626/6 = .438$. Similarly, the proportion of the total variance extracted by F_2 is .409. These proportions are shown in the last row of Table 19.4. The proportion of the total original variance extracted by the two factors together is .847 (or 84.7%). This result is another indication that the factor analysis has been quite effective.[9]

Let us now turn to interpretation of F_1 and F_2. The first factor is strongly correlated with X_1, X_4, and X_6 and weakly correlated with X_2, X_3, and X_5. The second factor is strongly correlated with X_2, X_3, and X_5

---

[9] A factor analysis that attempts to capture *all* the variance in the original variables is sometimes called *principal-components analysis.* A variation of this procedure—called *common-factor analysis*—attempts to capture only the part of the total variance that is shared by all the original variables. For further discussion of the distinction between the two procedures, see, for example, Joseph F. Hair, Jr., Rolph E. Anderson, and Ronald L. Tatham, *Multivariate Data Analysis* (New York: Macmillan, 1987), pp. 240–241.

and weakly correlated with X_1, X_4, and X_6. These results are somewhat fortunate; factor loadings will not always be so clean as to suggest that key variables making up one factor have little in common with another factor.[10] What do X_1, X_4, and X_6 have in common? Not minding the product's high price (X_1), impressing friends with the product (X_4), and viewing the product as being vastly superior to competing products (X_6) suggest that the common dimension underlying them can be called a *prestige* factor. Similarly, satisfaction with the products' capabilities (X_2), lack of worry about the product failing (X_3), and believing that the product is technologically advanced (X_5) imply that F_2 can perhaps be labeled as a *performance* factor. However, as we pointed out earlier, these labels are subjective and may be challenged by others as being inappropriate.

Not all factor loadings have to be positive, as in Table 19.4. After all, they are correlation coefficients and can range from -1 to $+1$. The features the original variables are measuring and the nature of the interrelationships among them will have an impact on the signs of the factor loadings. For instance, if a statement such as "Star VCRs are technologically behind the times" had been included in the SBI survey, its loading on factor F_2 in Table 19.4 would most likely have been negative. Intuitively, a negative factor loading means that the variable is representing something that is contrary to what the overall factor dimension represents.

Potential Applications

Factor analysis has several different applications. First, it can be used in developing parsimonious, yet comprehensive, multiple-item scales for measuring various marketing constructs. This application is perhaps the most frequent one, judging from published research reports and articles. In Chapter 11 we saw that the process of developing a multiple-item scale (e.g., a Likert or semantic-differential scale for measuring attitudes) typically starts with generating a large set of statements related to a topic. Factor analysis can be helpful in reducing the set of statements to a concise instrument and, at the same time, ensuring that the retained statements adequately reflect critical aspects of the construct being measured. The following example illustrates this application.

Example. A public utility company wants to develop a 15-item scale to measure the attitudes of its customers toward nuclear power. The company has already generated an initial pool of 100 items concerning nuclear power. Further, it has data on these items in the form of ratings obtained through a pilot survey of customers.

---

[10] When the initial factor-loading matrix is difficult to interpret, one can sometimes make the pattern of loadings clearer through a process called *factor rotation*. Details of factor rotation are complex and can be found in advanced texts, such as Nunnally, *Psychometric Theory*, and Harman, *Modern Factor Analysis*.

One approach to constructing the final 15-item scale is to use factor analysis in the following fashion: As was stated in Chapter 11, attitudes are generally believed to have three distinct dimensions, cognitive, affective, and behavioral. One can therefore conduct a factor analysis of the initial pool of 100 items and extract three factors from it. An examination of the resulting factor-loading matrix and the key items making up each factor can indicate whether the three attitude dimensions are adequately covered by the initial set of items. If they are, the five items with the highest loadings on each factor can be chosen for the final 15-item scale.[11]

A second application of factor analysis is to bring into bold relief the nature of distinct dimensions underlying an existing data set and hence offer managerial insights that may not emerge otherwise—insights that can be helpful in developing market segmentation and marketing-mix strategies. For instance, the factor-loading matrix pertaining to the SBI survey data (Table 19.4) implies that prestige and performance may be two critical criteria underlying customer evaluations of Star VCRs. This finding, in turn, suggests that SBI may benefit by building its promotional strategy for Star VCRs around the themes of prestige and performance.

Third, the ability of factor analysis to convert a large volume of data into a set of factor scores on a limited number of *uncorrelated* factors makes the technique ideal for use in conjunction with other analysis procedures such as multiple regression and discriminant analysis. Suppose 20 closely related independent variables are to be included in a multiple regression equation. Two potential problems here are multicollinearity and the lack of a sufficient number of observations. An effective way of circumventing both problems is to perform a factor analysis of the 20 independent variables. Since the variables are closely related, just a few factors, say 4, should capture most of the information in the independent-variable data set. The 4 uncorrelated factors can be treated as independent variables, and their factor scores can be used as raw data for the multiple regression analysis.[12]

---

[11] A number of applications of factor analysis in scale development have been reported in the marketing literature; see, for example, Jehoshua Eliashberg and Thomas S. Robertson, "New Product Preannouncing Behavior: A Market Signaling Study," *Journal of Marketing Research,* 25 (August 1988), pp. 282–292.

[12] Despite the versatility and mathematical sophistication of factor analysis, subjective decisions made by users of the technique play a critical role in how meaningful its applications are. For a discussion of these decisions and their impact on the final outcomes, see William D. Wells and Jagdish N. Sheth, "Factor Analysis," in *Handbook of Marketing Research,* ed. Robert Ferber (New York: McGraw-Hill, 1974), pp. 2.248–2.471; see also David A. Aaker, "Factor Analysis: An Exposition," in *Multivariate Analysis in Marketing,* ed. David A. Aaker (Palo Alto, Calif.: Scientific Press, 1981), pp. 163–171.

CLUSTER ANALYSIS

Cluster Analysis
A technique for segmenting objects (e.g., customers, market areas, products) into homogeneous groups, given data for the objects on a variety of characteristics.

As the name of this technique implies, the basic purpose of ***cluster analysis*** is to segment objects (e.g., customers, market areas, and products) into groups so that objects within each group are similar to one another on a variety of characteristics. Cluster analysis is potentially valuable in market segmentation studies in which the objective is to identify distinct customer groups. This technique strives to identify *natural groupings* of objects on the basis of their values on a number of variables without designating any of them as a dependent variable. The following example illustrates this point.

Example. A firm offering recreational services wants to enter a new region of the country. It recently surveyed a large sample of households in this region and gathered data on over a hundred characteristics including demographics, expenditures on recreation, leisure time activities, and interests of household members. The firm wants to identify one or several household segments that are likely to be most responsive to its advertising and to its services.

One way for the firm to identify such segments is to conduct a cluster analysis of the data it has gathered. The results of this analysis will reveal clusters of households where each cluster contains households that have similar data on the measured characteristics and where each cluster differs markedly from other clusters. Examining the composition of individual clusters will aid the firm in deciding which clusters to target and how best to reach them through its advertising.[13]

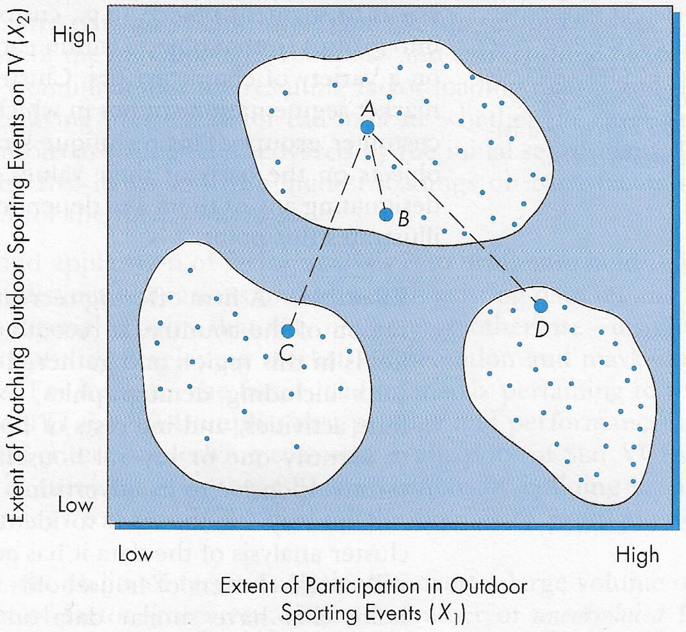
It is important to note that cluster analysis is different from discriminant analysis even though both techniques deal with partitioning a collection of objects into groups so that each group contains similar objects. In discriminant analysis the groups are *prespecified* by the researcher on the basis of a dependent variable. The technique then examines the ability of a set of independent variables to distinguish effectively between the various groups. In cluster analysis the groups are not prespecified. Instead, the technique itself generates the groups on the basis of similarity of the objects on a number of variables.

How does cluster analysis work? Several clustering procedures are available, each based on a somewhat different set of complex computer

---

[13] Several marketing research firms specialize in performing cluster analysis to identify useful customer segments for their clients. For a discussion of several such firms and the services they offer, see Doris Walsh, "Cluster Demographics Can Target Consumers but Zip-Code Groupings May Not Be Accurate," *Marketing News,* May 13, 1983, Section 1, pp. 24–25.

Exhibit 19.8 *Clusters Formed by Using Data on Two Characteristics*



routines.[14] But the basic principle underlying them is the same and involves measuring the similarity between objects on the basis of their values on the various characteristics. Similarity between objects is often ascertained through some *distance measure,* which can best be illustrated in the context of a situation involving just two clustering characteristics. Suppose data are available for a sample of individuals on the two variables described below:

- The extent to which they participate in outdoor sporting events (X_1).
- The extent to which they watch outdoor sporting events on TV (X_2).

Data on X_1 and X_2 are plotted on a two-dimensional graph in Exhibit 19.8.

---

[14] For additional details, see Ronald A. Frank and Paul E. Green, "Numerical Taxonomy in Marketing Analysis: A Review Article," *Journal of Marketing Research,* 5 (February 1968), pp. 83–98. Also, a variation of the factor analysis procedure we discussed can be used for clustering *objects* (instead of variables). This analysis is called Q-factor analysis, to distinguish it from traditional factor analysis, which is called R-factor analysis (see Hair, Anderson, and Tatham, *Multivariate Data Analysis,* pp. 239–240).

Each point in Exhibit 19.8 represents one individual. The physical distance between any pair of points is *inversely* related to how similar the corresponding individuals are when their X_1- and X_2-values are considered together. Thus individual A is more like B than either C or D. From a comparison of all such interpoint distances, Exhibit 19.8 reveals three distinct clusters.

Constructing clusters from a scatter plot—that is, determining how many distinct clusters there are and to which cluster each data point should be assigned—is an iterative, trial-and-error process. Computer algorithms are available for this purpose and must be used if the clustering is to be done in an efficient, systematic fashion. The mechanics underlying the computer algorithms are quite involved and vary across algorithms. However, the basic idea behind all of them is to start with some arbitrary cluster boundaries and modify them progressively until a stage is reached where the average interpoint distances within clusters are as small as possible relative to average distances between clusters. This logic of minimizing interpoint distances within clusters and maximizing them across clusters also applies to situations where a large number of variables form the basis for clustering.

The use of cluster analysis is not limited to segmenting customers. It can also be used to segment objects such as geographic market areas or brands within a product category. One application of cluster analysis, for instance, involved segmenting 88 standard metropolitan areas in the United States into homogeneous clusters on the basis of several economic and demographic characteristics of the areas. The results, which revealed 18 distinct clusters, can be valuable to marketers who want to select similar markets for conducting controlled research studies.[15] A variety of other marketing applications of cluster analysis are summarized in a review article on this topic.[16]

MULTIDIMENSIONAL SCALING

Multidimensional Scaling
A technique for inferring the number and the nature of dimensions underlying respondent evaluations/perceptions on the basis of similarity and/or preference judgments provided by respondents about objects (e.g., companies, products, brands).

Multidimensional scaling is most often used in marketing to identify the relative positions of competing brands, stores, and the like, as perceived by customers, and to uncover key dimensions underlying customers' evaluations. The technique is intended to infer the underlying dimensions from a series of similarity and/or preference judgments provided by customers about objects (products, brands, etc.) within a given set. In this respect multidimensional scaling does somewhat the reverse of what cluster analysis does: While cluster analysis groups objects according to similarities inferred from data on prespecified dimensions, multidimen-

---

[15] Paul E. Green, Ronald E. Frank, and Patrick J. Robinson, "Cluster Analysis in Test Market Selection," *Management Science,* April 1967, B–387 to B–400.

[16] Girish Punj and David W. Stewart, "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research,* 20 (May 1983), pp. 134–148.

Table 19.5 *Similarity Rankings of Six Department Stores*

	K-Mart	Penneys	Sears	Walmart	Wards	Woolworth
K-Mart		12	11	1	7	3
Penneys			5	15	4	10
Sears				13	6	14
Walmart					9	2
Wards						8

Note: Numbers are ranks indicating perceived similarities between pairs of stores; the smaller the number, the more similar the pair of stores is.

sional scaling infers underlying evaluative dimensions from similarities and/or preferences indicated by customers. Data on perceived similarities or preferences can be in the form of ranks (i.e., nonmetric) or in the form of more refined ratings (i.e., metric). Multidimensional-scaling approaches are available for analyzing nonmetric as well as metric input data; one may also analyze data on similarities and preferences separately or on both simultaneously.[17] To get an intuitive idea of what multidimensional scaling does, we will use a simple example involving analysis of similarities provided in the form of ranks.

Example. A customer is given a set of six department stores and asked to express how similar each store is to the others. Specifically, the customer is asked to compare *pairs* of stores and rank the pairs from most similar to least similar. Since there are six stores, 15 distinct pairs of stores are possible. The ranks given by the customer are shown in Table 19.5.

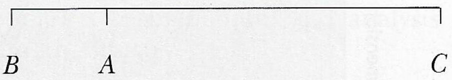
Multidimensional scaling, like cluster analysis, is an iterative process that can be carried out by using one of several available computer programs.[18] If the data in Table 19.5 are subjected to multidimensional scaling, the technique will attempt to generate a geometric configuration of the stores such that distances between pairs of stores are as consistent as possible with the customer's similarity ranks—that is, the pair of stores ranked 15th are farthest apart, the pair of stores ranked 14th are next farthest apart, and so on.

---

[17] For further information and technical details concerning multidimensional scaling, see Paul E. Green and Frank J. Carmone, *Multidimensional Scaling* (Boston: Allyn and Bacon, 1970); Paul E. Green, "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing*, 39 (January 1975), pp. 24–31.

[18] For a description of these programs, see Paul E. Green and Vithala R. Rao, *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms* (New York: Holt, Rinehart and Winston, 1972).

An important feature of this approach is that it attempts to represent objects in a geometric space with the *least* possible number of dimensions necessary to make the resulting configuration consistent with the similarity ranks. For instance, suppose three objects, *A, B,* and *C*, are such that a respondent perceives *A* and *B* to be most similar (rank = 1), *B* and *C* to be least similar (rank = 3), and *A* and *C* to be somewhere in between (rank = 2). These similarity rankings can be easily transformed into a corresponding geometric configuration in just *one* dimension, as follows:



However, as the number of objects increases, a higher-dimensional space may be necessary to represent the objects so that interobject distances are consistent with perceived similarity rankings.

Conceptually, this technique starts by arraying objects on a straight line (one dimension). If the object cannot be positioned on this line in accordance with perceived similarities between them, it adds another perpendicular line and moves the objects around in two-dimensional space to see whether a configuration that is reasonably consistent with the similarity rankings can be obtained.[19] If it cannot be obtained, the technique adds a third dimension, reconfigures the objects in three-dimensional space, and so on. Exhibit 19.9 shows a two-dimensional configuration of the six stores in which the interstore distances are consistent with the input ranking shown in Table 19.5.

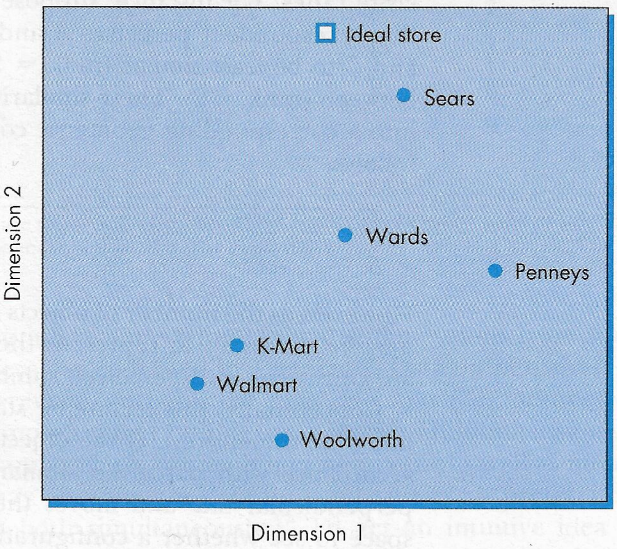
Ignoring for the moment the point marked Ideal store, what insight does Exhibit 19.9 offer? To answer this question, we must know what the two dimensions represent. Labeling the dimensions is quite subjective, however, and is hence a potential problem area. It basically involves inspecting the relative position of the objects along each dimension and inferring what the dimension is most likely to represent on the basis of one's prior knowledge about the objects themselves.

When the six stores in our example are arrayed along the horizontal dimension (dimension 1), Walmart is closest to the origin and Penneys is farthest from the origin. From what we know about the stores, we can perhaps label dimension 1 as price, with prices increasing from left to right along the dimension. Likewise, dimension 2 can be interpreted as representing number of product lines or product variety, with variety increasing from bottom to top. (Remember, however, that someone else looking at the same map may come up with a different set of labels for

---

[19] What is reasonably consistent is based on a measure called the *stress index,* whose value is provided by most multidimensional-scaling programs after each iteration. For details about the stress index, see J. B. Kruskal, "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika,* March 1964, pp. 1–27.

Exhibit 19.9 *Multidimensional Map of Department Stores Based on Similarity Rankings*



the two dimensions. Also, in some cases interpreting certain dimensions meaningfully may be impossible.)

Thus from Exhibit 19.9, which summarizes the results of applying multidimensional scaling to the customer's similarity rankings, we can infer that this particular customer implicitly used price and product variety as key criteria in comparing the six stores. These criteria may therefore be critical in influencing the customer's choice of department stores.

Now let us look at the position marked Ideal store in Exhibit 19.9. As its name suggests, this label signifies a hypothetical store most able to meet the customer's specific needs. The position of the ideal store can be derived by including it in the set of stores initially compared by the customer—in other words, by asking the customer to make pairwise comparisons among *seven* (six actual and one ideal) stores. From the position of the ideal store in Exhibit 19.9, Sears would appear to be most attractive and Woolworth least attractive to the customer. Insights like these will suggest changes in strategy and the dimensions along which such changes must be made in order to move a particular store closer to a customer's ideal store.

The store configuration shown in Exhibit 19.9 is derived from just one customer's perceived similarities among the stores and hence does not necessarily reflect the perceptions of other customers. For the same set of stores another customer's similarity rankings may produce an entirely different multidimensional map. Herein lies another practical limitation of multidimensional scaling (in addition to the problem of naming the dimensions): One may not be able to derive global inferences from the results of multidimensional scaling when the maps corresponding to different customers vary greatly, especially with respect to the number of dimensions in each. In such a situation one can attempt to identify *segments* of customers with fairly similar multidimensional maps by using an appropriate cluster analysis technique in conjunction with multidimensional scaling.[20] When properly conducted, multidimensional scaling can be helpful in areas such as product or brand positioning, market segmentation, and analysis of competing firms' positions within an industry.

CONJOINT ANALYSIS

Conjoint Analysis *A technique for deriving the utility values that respondents presumably attach to various attributes of an object on the basis of respondents' overall preferences for different bundles of attributes or descriptive profiles of the object.*

Conjoint analysis (also called conjoint measurement) is a technique for deriving the utility values that customers presumably attach to different levels of an object's attributes. Conjoint analysis requires respondents to compare hypothetical products, brands, and so on. The hypothetical stimuli are descriptive profiles formed by systematically combining varying levels of certain key attributes.[21]

Let us now discuss briefly how conjoint analysis works.[22] The first step in performing conjoint analysis is to construct several descriptive profiles (i.e., hypothetical stimuli) by combining different levels of selected attributes. Suppose we want to assess the role played by the following attributes in customer evaluations of personal computers: price, memory (data storage capacity), and versatility (number of different tasks the computer can perform). Also, suppose we are particularly interested in evaluating three levels of price ($1000, $1500, and $2000), two levels of memory (64K and 128K), and four levels of versatility (4, 8, 12, and 16 different tasks can be performed). With these levels for the three attributes a total of 24 different descriptive profiles of personal computers are possible (3 levels of price × 2 levels of memory × 4 levels of versatility).

---

[20] See, for example, Yoram Wind and Patrick J. Robinson, "Product Positioning: An Application of Multidimensional Scaling," in *Attitude Research in Transition,* ed. Russell I. Haley (Chicago: American Marketing Association, 1972), pp. 155–175.

[21] Like input data in multidimensional scaling, the input data in conjoint analysis can have one of several different formats. The most frequently used type of data is a set of preference rankings in conjoint analysis and similarity rankings in multidimensional scaling. For additional details, see Paul E. Green and V. Srinivasan, "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research,* September 1978, pp. 103–123.

[22] An excellent article that discusses conjoint analysis in very readable terms is Paul E. Green and Yoram Wind, "New Way to Measure Consumers' Judgements," *Harvard Business Review,* July/August 1975, pp. 107–117.

Two-Factors-at-a-Time (Trade-Off) Approach *A data-gathering procedure used in conjoint analysis, wherein respondents rank their preferences for various combinations of attribute levels, considering just a pair of attributes at a time.*

Full-Profile Approach *A data-gathering procedure used in conjoint analysis, wherein respondents rank-order their preferences for complete stimuli, each of which is a different combination of levels of all attributes.*

Utility Value *A number (obtained through conjoint analysis) representing a respondent's degree of preference for a certain level of a given attribute.*

One approach to obtaining customer-preference-ranking data suitable for conjoint analysis is called the ***two-factors-at-a-time***, or ***trade-off, approach***. In this approach customers are asked to rank their preferences for various combinations of attribute levels, considering just a pair of attributes at a time. Considering price and memory in our example, customers are asked to rank the six possible combinations of levels according to their preferences (most preferred = 1 and least preferred = 6). Similar sets of preference ranks can be obtained for the price-versatility and memory-versatility attribute pairs. Another way of obtaining customer-preference-ranking data is to use a ***full-profile approach***. In this approach customers are asked to rank-order their preferences for the 24 different profiles representing all possible combinations of the three attributes. For each approach, Research in Use 19.2 presents illustrations of instructions and stimuli that are typically used in gathering data for conjoint analysis.

The two-factors-at-a-time approach is easier for customers to comprehend and respond to, especially when numerous attributes are included in the analysis. This approach is somewhat unrealistic, however, because it involves presenting only partial stimuli to respondents. Hence there is some question about how meaningful the preference ranks will be. The full-profile approach is more realistic but may overwhelm respondents when the number of attributes is large.[23]

The output of conjoint analysis is a set of ***utility values*** corresponding to each attribute. The higher the utility value for a certain level of an attribute, the greater is a customer's preference for descriptive profiles containing that level. For instance, assume that the utility values plotted in Exhibit 19.10 are generated by a conjoint analysis of preference rankings given by one customer. This exhibit offers several interesting insights. As we would expect, the utility of price decreases for the higher price levels, while the utilities of memory and versatility go up with increases in their attribute levels. The relative importance of the three attributes can be ascertained by computing and comparing their respective *ranges* of utility values:

$$\begin{aligned}\text{range for price} &= 0.8 - 0.3 = 0.5,\\ \text{range for memory} &= 0.9 - 0.6 = 0.3,\\ \text{range for versatility} &= 0.8 - 0.4 = 0.4.\end{aligned}$$

Intuitively, a wide range of utility values should signify that customer preferences are quite sensitive to changes in the levels of the attribute. Hence in our illustration price is most critical, versatility is next most critical, and memory is least critical in influencing customer preferences for personal computers.

---

[23] Further discussion of the two approaches is given in Green and Srinivasan, "Conjoint Analysis in Consumer Research."

Research in Use 19.2 *Personal-Computer Study*

Instructions and Stimuli for Data Collection in Conjoint Analysis

TWO-FACTORS-AT-A-TIME APPROACH

We would like to know what your relative preferences are concerning personal-computer features and how much of one feature you are willing to trade off against another feature. The following table has six blank boxes representing different combinations of price and memory of personal computers. Please indicate your relative preferences for the different combinations by simply writing the number 1 in the box that represents your *most preferred* price-memory combination, then writing the number 2 in the box that represents your *second most preferred* price-memory combination, and so on, until you have assigned the number 6 to your least preferred combination.

	Price		
Memory	**$1,000**	**$1500**	**$2000**
64K			
128K			

Note: This rating procedure would be repeated for the 12 price-versatility combinations and the 8 memory-versatility combinations.

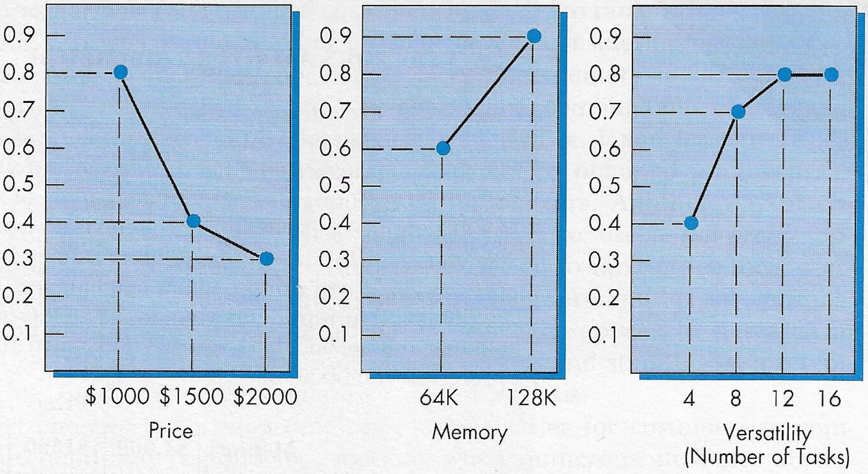
Full-Profile Approach

Here are 24 cards representing different combinations of price, memory, and versatility levels of personal computers. Please study each card and place it in one of three categories—(a) definitely prefer; (b) indifferent; (c) definitely not prefer—according to how much you prefer a personal computer that has the features listed on the card. After you have allocated all 24 cards across the three categories, rank-order the cards *within each category* from "most preferred" to "least preferred."

PERSONAL COMPUTER	PERSONAL COMPUTER	PERSONAL COMPUTER	PERSONAL COMPUTER
Price $1000	Price $1000	Price $1000	Price $1000
Memory 64K	Memory 64K	Memory 64K	Memory 64K
Versatility 4 tasks	Versatility 8 tasks	Versatility 12 tasks	Versatility 16 tasks

Note: Only 4 of the 24 cards are shown. Also, the stack of cards will typically be shuffled—to randomize the order—before being given to each respondent.

Exhibit 19.10 *Utility Values for Three Personal-Computer Attributes*



The utility values in Exhibit 19.10 can also be used to compare the potential attractiveness of different personal-computer configurations. For instance, is a personal computer with a 64K memory, capable of performing 8 different tasks, and costing $1500 likely to be more attractive to the customer than one with a 128K memory, capable of performing 12 different tasks, and costing $2000? This question can be answered by computing the total utility of the two hypothetical product configurations, as follows:[24]

total utility for the 64K, 8-task, $1500 personal computer
$= 0.6 + 0.7 + 0.4$
$= 1.7,$

total utility for the 128K, 12-task, $2000 personal computer
$= 0.9 + 0.8 + 0.3$
$= 2.0.$

Thus the latter combination of attributes is likely to be more attractive to the customer than the former.

---

[24] This computation assumes that a simple, linear, additive conjoint model underlies the customer preference rankings. However, other conjoint models that may produce somewhat different results are also available. For a comparison of such models, see Ishmael Akaah and Pradeep K. Korgaonkar, "An Empirical Comparison of the Predictive Validity of Self-Explicated, Huber-Hybrid, Traditional Conjoint, and Hybrid Conjoint Models," *Journal of Marketing Research,* 20 (May 1983), pp. 187–197; see also Paul E. Green, Abba M. Krieger, and Pradeep Bansal, "Completely Unacceptable Levels in Conjoint Analysis: A Cautionary Note," *Journal of Marketing Research,* 25 (August 1988), pp. 293–300.

Like the results in multidimensional scaling, the results in conjoint analysis are individual-specific. That is, a set of utility values are generated for *each* respondent. Utility values must therefore be pooled across respondents if generalizable inferences are to be made. Although pooling of utility values can be troublesome, several meaningful approaches have been developed for this purpose.[25] Some of these approaches can yield pooled utilities and also aid in estimating market shares for products with different combinations of attribute levels. For example, AT&T used conjoint analysis in conjunction with a computer simulator to identify the most attractive combination of attributes for a new data terminal and to predict its market share. Actual sales of this product after it was introduced showed that the market share prediction was very accurate.[26] A recent survey found that conjoint analysis has had several hundred commercial applications within the past few years.[27] The practical appeal of conjoint analysis results and the ready access to high-speed computers are likely to further increase the use of the technique in the future. Even some of the data collection problems associated with conjoint analysis—such as the difficulty in systematically formulating descriptive profiles when the number of attributes are numerous and in presenting the profiles without overwhelming the respondents—are being overcome by using computers to construct the stimuli, present them to respondents, record their preferences, and transfer the data immediately to a central computer's memory.[28]

SUMMARY

This chapter discussed a variety of multivariate techniques, broadly classified into dependence and interdependence techniques. The primary focus of the chapter was on providing a conceptual understanding of the techniques along with their potential applications and limitations. The dependence techniques covered in this chapter included analysis of variance (ANOVA) and discriminant analysis, while the interdependence techniques covered included factor analysis, cluster analysis, multidimensional scaling, and conjoint analysis.

The seeming statistical sophistication of multivariate techniques, coupled with the ready access to computer programs incorporating the techniques, may tempt researchers to rush into using them without adequately considering their appropriateness in a given situation. Yielding to the temptation may lead to meaningless results, however.[29] One must also bear in mind that applications of multivariate techniques, despite their apparent quantitative rigor, may not be totally free of researcher subjectivity. For instance, it is up to the researcher to decide which independent variables to include in any application involving the use of dependence techniques. Yet the choice of a particular set of independent variables can greatly influence quantitative results and the inferences they imply.

---

[25] See, for example, William L. Moore, "Levels of Aggregation in Conjoint Analysis: An Empirical Comparison," *Journal of Marketing Research,* 17 (November 1980), pp. 516–523; see also Paul E. Green, "Hybrid Models for Conjoint Analysis: An Expository Review," *Journal of Marketing Research,* 21 (May 1984), pp. 148–154.

[26] Attitude Research, Conjoint Analysis Guided Ma Bell's Entry into Data-Terminal Market," *Marketing News,* May 13, 1983, Section 1, p. 13.

[27] Philippe Cattin and Dick R. Wittink, "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing,* 46 (Summer 1982), pp. 44–53.

[28] An illustration involving computerized data collection is discussed in Linda D. Straube and Bridgid J. Michand, "Combine Microcomputer Interviewing with Conjoint Analysis to Study Pricing Strategies," *Marketing News,* May 13, 1983, Section 2, p. 10.

KEY TERMS

Dependence technique
Interdependence technique
Analysis of variance (ANOVA)
Treatment
Total sum of squares
Within sum of squares
Between sum of squares
Main effect
Completely randomized ANOVA
Blocking factor
Randomized-block ANOVA
Latin-square ANOVA
Factorial ANOVA
Interaction effect
Analysis of covariance (ANCOVA)
Covariate
Discriminant analysis
Discriminant function
Discriminant weight (coefficient)
Discriminant score
Discriminant axis
Confusion matrix
Hit rate
Proportional chance criterion
Maximum chance criterion
Factor analysis
Factor score
Factor loading
Factor-loading matrix
Achieved communality
Eigenvalue
Cluster analysis
Multidimensional scaling
Conjoint analysis
Two-factors-at-a-time (trade-off) approach
Full-profile approach
Utility value

QUESTIONS

1. A marketer of a refrigerated fruit drink has developed three new package designs for the product that will enable the fruit drink to be stored at room temperature. The marketer wants to know if any or all of these package designs may result in significantly higher sales than the current package design that requires refrigerated storage. Suggest a suitable research approach to the marketer for realistically evaluating the alternative package designs. What technique(s) should

---

[29] For a discussion of caveats to bear in mind while using multivariate techniques, see Jagdish N. Sheth, "Seven Commandments for Users of Multivariate Methods," in *Multivariate Methods,* ed. Jagdish N. Sheth (Chicago: American Marketing Association, 1977), pp. 333–335.

be used to analyze the research data to fulfill the marketer's information needs? Explain the rationale for your answer.

2. Explain in simple words the distinction between a randomized-block ANOVA design and an ANCOVA design.

3. When is "interaction effect" said to occur? Illustrate this effect with a suitable example.

4. A random sample of 600 respondents consisted of 300 users (group 1) and 300 nonusers (group 2) of a certain product. The means and standard deviations of the age (variable *A*) and income (variable *B*) of the respondents in the two groups are as follows:

Group 1 $\bar{A}_1 = 30$ $\bar{B}_1 = \$15{,}000$ $S_{1A} = 3$ $S_{1B} = \$100$

Group 2 $\bar{A}_2 = 40$ $\bar{B}_2 = \$20{,}000$ $S_{2A} = 4$ $S_{2B} = \$90$

A discriminant analysis of the age and income data for the two groups produced the following discriminant function:

$$Y = v_A A + v_B B \quad \text{where} \quad v_A = .10, \quad v_B = .01.$$

a. From the discriminant analysis, would you classify a person aged 25 with an income of $18,000 as a user or a nonuser? Why?

b. Is one of the two independent variables more important than the other in discriminating between users and nonusers? If so, what exactly is the relative importance of one variable with respect to the other? If not, why not?

5. Describe in your own words what *factor analysis* is, and indicate the potential benefits it can offer.

6. Consider the following factor-loading matrix:

	F_1	F_2	F_3
X_1	.90	.02	.05
X_2	.80	.02	.10
X_3	.05	.90	.02
X_4	.03	.03	.95
X_5	.75	.05	.10
X_6	.10	.80	.05
X_7	.60	.06	.08
X_8	.05	.85	.03

a. What proportion of the total variance in the original set of variables is explained by the three factors taken together?

b. What are the achieved communalities for X_4 and X_7? How would you interpret these values?

c. Are the three factors easy to interpret? Why or why not?

7. Exhibit 18.9 in Chapter 18 contains the correlation matrix for three variables: sales revenue (SALES), advertising expenditures (ADVTG), and number of competing brands (NCOMP). Suppose the data on these three variables are subjected to factor analysis. Will one factor be able to account for a substantial portion of the variance in the original data set? Explain your answer. (If you have access to a computer, subject the raw data in Table 18.7 to factor analysis and see whether the results agree with your answer).

8. Refer to Question 7. Assuming that one factor is able to capture the essence of SALES, ADVTG, and NCOMP, what label will you place on that factor?

9. State how the following techniques differ from each other: (a) cluster analysis; (b) discriminant analysis.

10. Define *conjoint analysis* and state its potential application.

11. "Multivariate techniques, though mathematically sophisticated, are not free of researcher and/or decision maker subjectivity." Defend this statement by pointing out some subjective aspect of each technique we discussed.

12. A bank conducted a survey of customers and obtained preference rankings for various hypothetical bank profiles constructed by using the following characteristics and levels within each characteristic:

Operating hours:
A. 8 a.m. to 3 p.m. weekdays
B. 9 a.m. to 5 p.m. weekdays
C. 9 a.m. to 5 p.m. weekdays plus 9 a.m. to noon Saturdays

Automatic-teller machines (to facilitate routine transactions 24 hours every day):
A. Yes B. No

Monthly service charge:
A. None C. $15
B. $5 D. $20

Conjoint analysis of the data yielded the following utility values for various levels of each of the three characteristics:

- Operating hours: A = 0.3; B = 0.3; C = 0.8.
- Automatic-teller machines: A = 0.6; B = 0.3.
- Monthly service charge: A = 1.0; B = 0.8; C = 0.4; D = 0.1.

a. Which of the three bank characteristics is most crucial in influencing customer preference? Why?
b. Will the bank benefit more by operating 8 a.m. to 3 p.m. weekdays, providing no automatic-teller machines, and assessing no monthly service charge, or by operating 9 a.m. to 5 p.m. weekdays plus 9 a.m. to noon Saturdays, providing automatic-teller machines, and assessing a monthly service charge of $15? What if the service charge is changed to $20 in the latter alternative?

APPENDIX TO CHAPTER 19

ANOVA and Dummy-Variable Regression

Dummy Variable
A variable that takes on a value of either 1 or 0. Categorical variables that normally would not be included in certain analysis procedures (e.g., regression analysis) can be included in such procedures after converting the variables into a set of dummy variables. A categorical variable with k categories will require k−1 dummy variables to represent it.

The purpose of this appendix is to demonstrate the conceptual correspondence between ANOVA and regression analysis and also to discuss ***dummy variables***, which are necessary when data on one or more independent variables in a regression equation are categorical rather than continuous. Examples of categorical independent variables frequently encountered in marketing research studies include sex, ethnic background, marital status, and occupation.

Both ANOVA and regression analysis are meant to shed light on the influence of independent variables on some dependent variable. We have already seen certain computational similarities between the two techniques (e.g., the analogous total as well as partitioned sums of squares). We can gain a more thorough understanding of the equivalence between the two techniques by performing a regression analysis of our illustrative data in Table 19.2 presented in the chapter, using dummy independent variables. Before doing so, however, we need to know more about the concept of a dummy variable.

A dummy variable is defined as one that can only have a value of either 0 or 1. While independent variables in regression analysis are typically required to be at least interval-scaled, an exception to this general requirement is a 0–1 dummy variable. For reasons that we will not get into here, it is legitimate to include categorical independent variables in a regression equation as long as data on them can be coded as 0s and 1s. Suppose, for example, we want to include sex as an independent variable in a multiple regression equation for predicting the extent of use of a certain product. Since sex has only two categories, it naturally fits the description of a dummy variable. So we can include, in the regression equation, an independent variable, say S, coded as

$S = 0$ if respondent is male,
$S = 1$ if respondent is female.

What if a categorical independent variable has more than two categories, such as TREAT in our ANOVA example? We can still include it in a regression equa-

Table A.19.1 *Modified Data Set with Dummy Independent Variables*

		Independent Variables	
Store (i)	Dependent Variable, Sales (Y_i)	DUMV1 (X_{1i})	DUMV2 (X_{2i})
1	37	0	0
2	38	0	0
3	40	0	0
4	40	0	0
5	38	0	0
6	38	0	0
7	40	0	0
8	39	0	0
9	46	1	0
10	43	1	0
11	43	1	0
12	45	1	0
13	45	1	0
14	43	1	0
15	44	1	0
16	44	1	0
17	46	0	1
18	49	0	1
19	48	0	1
20	48	0	1
21	47	0	1
22	48	0	1
23	49	0	1
24	49	0	1

tion by creating a suitable *set* of dummy variables. Specifically, any variable with k categories can be recoded into a 0–1 format by forming $k - 1$ dummy variables. To illustrate, the three-category TREAT variable can be represented by two dummy variables, say DUMV1 and DUMV2, as follows:

TREAT	DUMV1	DUMV2
1. Regular price	0	0
2. 50¢ off	1	0
3. 75¢ off	0	1

Any sample store selling the product at regular price is equivalent to a store whose values on DUMV1 and DUMV2 are zero. Stores with the 50¢-off pricing treatment are uniquely identified by the combination DUMV1 = 1 and DUMV2 = 0. Similarly, stores with DUMV1 = 0 and DUMV2 = 1 are those with the 75¢-off pricing treatment. What we have done here is simply convert TREAT (coded as 1, 2, or 3) into two independent dummy variables in such a way that the three possible combinations of 0–1 values on them are uniquely associated with the three pricing treatment categories. A data set can now be created from the raw data in Table 19.2 in order to derive a multiple regression

Exhibit A19.1 *SPSS Computer Output for Analysis of the Data in Table 19.1, Using Dummy-Variable Regression*

```
                                   MULTIPLE REGRESSION

DEPENDENT VARIABLE. .       SALES                                                    VARIABLE LIST 1
VARIABLE(S) ENTERED ON STEP NUMBER 1. .      DUMV1                                 REGRESSION LIST 1
                                             DUMV2

MULTIPLE R           0.96386   ANALYSIS OF VARIANCE   DF    SUM OF SQUARES    MEAN SQUARE        F
R SQUARE             0.92903   REGRESSION             2.       345.25000        172.62500    137.44550
ADJUSTED R SQUARE    0.92227   RESIDUAL              21.        26.37500          1.25595
STANDARD ERROR       1.12069

              VARIABLES IN THE EQUATION                      VARIABLES NOT IN THE EQUATION

VARIABLE           B           BETA      STD ERROR B      F      VARIABLE  BETA IN  PARTIAL  TOLERANCE  F
DUMV1         5.375000      0.64391        0.56035     92.012
DUMV2         9.250000      1.10812        0.56035    272.502
(CONSTANT)    38.75000
ALL VARIABLES ARE IN THE EQUATION
STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.
```

equation of the following form:

$$\hat{Y}_i = a + b_1X_{1i} + b_2X_{2i},$$

where $\hat{Y}_i$ is the predicted sales for any store i in the sample of 24 stores, X_{1i} is the DUMV1 value for store i, and X_{2i} is the DUMV2 value for store i. The data set suitable for regression analysis is shown in Table A.19.1.

The computer output stemming from regression analysis of the data set in Table A.19.1 (using SPSS) is shown in Exhibit A.19.1.

The right-hand side of Exhibit A.19.1 contains basically the same information as the ANOVA table shown in Exhibit 19.1 presented in the chapter. In fact, the numbers are identical except for a few minor discrepancies due to rounding. The *F*-value for testing the significance of R^2 in the regression analysis output is the same as the *F*-value for testing the significance of the treatment effects in the ANOVA table. The two procedures are thus statistically equivalent and lead to the same conclusion: There is a significant association between price and unit sales of the store brand of vegetable oil.

The interpretation of the regression coefficient (*b*-value) of a dummy variable is somewhat different from the traditional interpretation of slope—namely, that it is the change in the dependent variable per unit change in the independent variable, assuming all other variables remain the same. Since a dummy variable merely indicates the presence (1) or absence (0) of some attribute, the notion of a continuous unit change in the attribute is meaningless. Moreover, when several dummy variables are used to represent one independent variable (as in our example), the particular *combination* of 0s and 1s corresponding to each independent-variable category must be considered in interpreting the dummy-variable regression coefficients. In the grocery store chain illustration a 0–0 combination (i.e., DUMV1 = 0 and DUMV2 = 0) represents the regular-price treatment. This combination can be called the *base level* of the independent variable. In general, the base level is one for which *all* dummy variables representing the independent variable have a value of 0. It is the benchmark against which the effects of other levels are measured by examining the regression coefficients. Specifically, the regression coefficient of any dummy variable has the following interpretation:

> Other things remaining the same, it represents the difference between the average dependent-variable value for units with a 1 on the dummy variable and the average dependent-variable value for the base level units.

To illustrate, the 1–0 combination of dummy-variable values (i.e., DUMV1 = 1 and DUMV2 = 0) represents stores receiving the 50¢-off pricing treatment. The regression coefficient for DUMV1 is 5.375 (see Exhibit A.19.1). This value represents the difference between the average unit sales of stores selling the vegetable oil at 50¢ off and the average unit sales of those selling it at regular price (base level):

average sales of stores selling at 50¢ off = 44.13

(see Table 19.2 presented in the chapter).

average sales of stores selling at regular price = 38.75

(see Table 19.2 presented in the chapter).

difference in average sales = 44.13 − 38.75 = 5.38,

which is the same as the *b*-value for DUMV1. The *b*-value of 9.25 for DUMV2 has a similar interpretation: Stores with a 1 on DUMV2 (i.e., those receiving the 75¢-off pricing tretament) on the average sell 9.25 units more than stores at the base level.

The statistical significance of the dummy-variable regression coefficients can be tested with the aid of the *F*-values reported in Exhibit A.19.1 and the testing procedure outlined in Chapter 18. It turns out that both coefficients are significant beyond the customary .05 significance level. Thus both the 50¢-off and the 75¢-off pricing treatments are significantly more effective in increasing unit sales than the regular-price treatment. In general, testing the statistical significance of dummy-variable regression coefficients is analogous to evaluating whether certain pairs of treatment effects are significantly different in ANOVA.[1]

The *a*-value (*Y*-intercept) in Exhibit A.19.1 is 38.75. That this value is the same as the average sales of base-level stores is not mere coincidence. In regression equations involving dummy as well as regular independent variables, the average dependent-variable value for base-level units (corresponding to a related set of dummy variables) will be part of the *Y*-intercept value. Since the regression equation in our example only has a pair of related dummy variables and no regular independent variables, the *Y*-intercept is identical to the average sales of base-level stores.

The foregoing discussion seems to suggest that categorical independent variables are not a constraint in conducting regression analysis since they can be readily converted into appropriate sets of dummy variables. A potential problem, however, is that too many dummy variables will sharply increase the number of observations and hence the sample size needed for the resulting multiple regression equation to be meaningful. Recall from Chapter 18 the rule of thumb calling for a minimum of 10 observations for every independent variable included in a regression equation. This rule of thumb applies to dummy variables as well.

---

[1] For a more formal discussion of the equivalence of these significance tests, see Fred N. Kerlinger and Elazur J. Pedhazur, *Multiple Regression in Behavioral Research* (New York: Holt, Rinehart and Winston, 1973), pp. 119–121.

Suppose, for instance, that three categorical independent variables, each with four levels, are to be included in a multiple regression equation after being converted to dummy variables. The number of dummy variables needed to represent the three variables in the regression equation is $(4 - 1) \times 3 = 9$. Unless the number of observations available is large enough to accommodate this many dummy variables, the regression equation will not be very useful. Furthermore, even when the sample size is sufficient, interpretation of the dummy-variable regression coefficients becomes more and more difficult as the number of categorical independent variables included in the regression increases. Thus dummy variables are not a panacea, and they must be used sparingly. Whenever feasible, collecting refined data on independent variables to begin with is better than collecting categorical data and conducting regression analysis with dummy independent variables.

QUESTIONS

1. What is the basic difference between ANOVA and regression analysis? In what ways can the two techniques be considered similar?
2. In analyzing data from a certain consumer survey using regression analysis, the researcher feels that race is a critical independent variable. Data on race were gathered by using the following categories: white, black, Hispanic, American Indian, Asian, and other. Can the researcher include race as an independent variable in the regression analysis? If so, explain how. If not, why not?

Cases for Part Six

CASE 6.1 *NORTHWEST BANK OF WOODCREEK (B)*

CASE 6.2 *ITT'S IMAGE ADVERTISING CAMPAIGN*

CASE 6.3 *RUBY RESTAURANT*

CASE 6.4 *HAMILTON BANK (B)*

CASE 6.5 *NATIONAL INSURANCE COMPANY*

CASE 6.6 *FANTASTIC CAR WASH*

CASE 6.1

Northwest Bank of Woodcreek (B)

The case titled "Northwest Bank of Woodcreek (A)" listed several specific objectives of the competitive study conducted for NBW by a local marketing research firm. One of those objectives was, "To examine the extent of recognition received by the slogans and symbols of the banks and savings and loan associations in Woodcreek." Two banks and three savings and loan associations in Woodcreek compete with NBW. They are Public Bank & Trust (PB&T), Woodcreek Savings Bank (WSB), People's Mutual Savings & Loan (PMS&L), First Federal Savings & Loan (FFS&L), and Home Savings & Loan (HS&L).

All three banks and three savings and loans in Woodcreek have their own distinct symbols that they use in their advertisements: PB&T uses the picture of an owl; WSB uses three triangles apparently forming the letter *W*; NBW uses a line drawing depicting a series of faces; PMS&L uses three pillars; FFS&L uses an eagle; and HS&L uses a fireplace. The three banks and one of the three savings and loans also have distinct promotional slogans: PB&T was "The wide awake bank"; WSB uses "The bank designed with you in mind"; NBW uses "Everybody's bank"; HS&L uses "There's no place like home."

SYMBOL AND SLOGAN RECOGNITION

During the personal interviews of area residents, the interviewers obtained recognition measures for the symbols and slogans used by the various institutions. For the purpose of measuring symbol recognition, each respondent was given a six-page booklet in which each page contained one of the six symbols. After looking at each symbol, the respondent was asked to state the name of the financial institution to which the

Table 1 *Recognition of Symbols of Institutions*

Symbol	Institution Using Symbol	Type of Respondents	Precent Saying “Don’t Know”	Percent of Respondents Attributing Symbol to								
				PB &T	WSB	NBW	PMS &L	FFS &L	PS &L	HS &L	MB &T	Other Institutions
Owl	PB&T	Customers	61.2	36.7	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
		Noncustomers	74.2	20.0	2.4	2.4	0.0	0.7	0.0	0.3	0.0	0.0
		All respondents	71.0	24.2	2.0	2.0	0.0	0.5	0.0	0.3	0.0	0.0
Three triangles	WSB	Customers	55.3	0.0	41.1	1.4	0.0	0.7	1.4	0.0	0.0	0.0
		Noncustomers	72.6	0.0	24.2	2.0	0.4	0.0	0.0	0.4	0.0	0.4
		All respondents	66.4	0.0	30.3	1.8	0.3	0.3	0.5	0.3	0.0	0.3
Faces	NBW	Customers	70.8	2.5	1.2	24.8	0.0	0.6	0.0	0.0	0.0	0.0
		Noncustomers	80.6	1.3	1.7	14.7	0.4	0.4	0.4	0.4	0.0	0.0
		All respondents	76.6	1.8	1.5	18.8	0.3	0.5	0.3	0.3	0.0	0.0
Three pillars	PMS&L	All respondents	83.3	0.3	2.3	1.5	9.2	0.8	2.0	0.3	0.8	0.0
Eagle	FFS&L	All respondents	83.7	0.8	1.0	2.8	0.0	11.5	0.0	0.0	0.0	0.3
Fireplace	HS&L	All respondents	16.5	0.3	0.3	0.8	0.0	0.8	0.0	81.4	0.0	0.0

Note: The percentages in rows labeled “All respondents” are based on a sample size of 393.

Table 2 *Recognition of Slogans of Institutions*

Slogan	Institution Using Slogan	Type of Respondents	Percent Saying “Don’t Know”	Percent of Respondents Attributing Slogan to								
				PB &T	WSB	NBW	PMS &L	FFS &L	PS &L	HS &L	MB &T	Other Institutions
“The wide awake bank”	PB&T	Customers	74.5	22.4	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
		Noncustomers	81.4	10.2	2.7	2.4	0.3	0.3	0.7	0.3	1.7	0.0
		All respondents	79.6	13.2	2.3	2.3	0.3	0.3	0.5	0.3	1.3	0.0
“The bank designed with you in mind”	WSB	Customers	51.1	3.5	30.5	10.6	0.7	1.4	0.0	0.7	0.7	0.7
		Noncustomers	52.8	3.6	21.4	17.1	0.0	0.0	0.8	3.8	0.0	1.2
		All respondents	52.2	3.6	24.7	14.8	0.3	0.5	0.5	2.3	0.3	1.0
“Everybody’s bank”	NBW	Customers	29.2	11.8	3.7	54.0	0.0	0.0	0.0	0.6	0.0	0.6
		Noncustomers	46.6	15.5	10.8	26.3	0.0	0.0	0.0	0.0	0.4	0.4
		All respondents	39.4	14.0	7.9	37.9	0.0	0.0	0.0	0.3	0.3	0.5
“There’s no place like home”	HS&L	All respondents	10.9	0.3	0.5	0.3	0.3	0.5	0.5	86.8	0.0	0.0

Note: The percentages in rows labeled “All respondents” are based on a sample size of 393.

symbol belonged. To measure slogan recognition, the interviewer simply read one slogan at a time and asked the respondents to name the institution that used the slogan.

RESULTS

The results of the symbol and slogan recognition tests are summarized in Tables 1 and 2, respectively. In both tables two additional financial institutions—Mercer Bank & Trust (MB&T) and Principal Savings and Loan (PS&L)—are included. These financial institutions are located in an adjoining community (outside Woodcreek city limits) to which some respondents incorrectly attributed several of the symbols/slogans used by the institutions in Woodcreek.

For the three Woodcreek banks the overall recognition measures are also broken out for customers and noncustomers of each bank. As an aid in interpreting the numbers in the tables, a couple of illustrations are provided here. In Table 2, for instance, in the row starting with the slogan "There's no place like home," the number in the fourth column is 10.9. This number indicates that 10.9% of all respondents could not attribute this slogan to any institution. The number in the seventh column of the same row indicates that 0.3% of all respondents incorrectly attributed this slogan to NBW. The number in the eleventh column shows that 86.8% of all respondents correctly attributed the slogan to HS&L. For each of the three Woodcreek banks the percentage columns have three numbers for each slogan. For instance, the row starting with "Everybody's bank" has the numbers 54.0, 26.3, and 37.9 in the seventh column. These numbers indicate that 54.0% of the customers of NBW, 26.3% of the noncustomers of NBW, and 37.9% of all respondents (customers and noncustomers combined) correctly attributed this slogan to NBW.

QUESTIONS

1. How do you interpret the effectiveness of the symbols and slogans of the Woodcreek area financial institutions on the basis of the results reported in Tables 1 and 2? What recommendation(s) would you make to the NBW management?
2. What analysis procedures do you think were used in arriving at Tables 1 and 2? What other procedures or tests, if any, would you recommend for a better interpretation of the findings?
3. Critically evaluate the clarity and meaningfulness of the two tables.

CASE 6.2

ITT's Image Advertising Campaign

The International Telephone and Telegraph Corporation (ITT), headquartered in New York, faced a serious image problem in the early seventies. A majority of the public was unaware of the firm and its variety of business activities. Furthermore, many of those who were aware of ITT had a rather blurred image of the firm, primarily because of confusion between the ITT and AT&T (American Telephone and Telegraph) names.

In 1974, ITT began a multiyear, multimillion-dollar advertising campaign aimed at building a distinct and favorable image for itself in the eyes of the public. The campaign involved a series of television and print ads that employed very creative approaches to convey memorable messages and slogans, such as "The best ideas are the ideas that help people." ITT's campaign won praise from advertising critics and received numerous awards and honors over the years, such as the CLIO and EFFIE awards. As of 1982, ITT was spending over $10 million on its corporate image-building campaign.

To assess the potential impact of the advertising campaign on the public, ITT retained the services of Yankelovich, Skelly and White, Inc., a marketing research firm well known for its expertise in conducting public opinion polls. Yankelovich, Skelly and White conducted a benchmark survey in 1974 to assess the public's views just before the start of the advertising campaign. The survey involved telephone interviews of a national sample of 1500 respondents. Similar surveys, using fresh random samples of respondents, were conducted at least once a year thereafter in order to monitor changes, if any, in the public's perceptions

---

The information in this case was provided by courtesy of International Telephone and Telegraph Corporation, 320 Park Avenue, New York, NY 10022.

Table 1 *ITT Corporate Strengths*

	Percent Response*						Percentage Point Difference	
Attribute	November 1978	October 1977	October 1976	December 1975	December 1974	January 1974†	October 1977	December 1975
One of the largest companies	85	86	90	86	85	68	−1	−1
Very profitable	77	76	80	73	76	61	+1	+4
Makes quality products	76	77	82	74	75	54	−1	+2
Leader in technology	76	75	81	73	72	49	+1	+3
Develops many new products	73	74	75	68	67	46	−1	+5
Good stock to buy or own	70	70	72	62	53	52	=	+8
Leads in research and development to improve products	68	69	71	66	65	46	−1	+2
Reliable	66	70	73	63	64	48	−4	+3
Gives good value	56	58	62	56	56	43	−2	=
Cares about its customers	51	58	59	55	56	40	−7	−4
Pays good wages	47	51	50	47	44	34	−4	=
Cares about general public	46	51	50	42	43	31	−5	+4
Fair to employees of all races	44	48	48	45	42	34	−4	−1
Good labor relations	43	42	46	40	43	29	+1	+3
Interesting place to work	43	42	45	40	40	26	+1	+3
Gives women equal opportunity	42	44	45	43	40	29	−2	−1
Cares about community near plant	40	40	44	40	40	26	=	=
Good balance in profits/public interest	40	35	37	33	31	29	+5	+7
Honest, forthright advertising	39	42	44	37	39	27	−3	+2
Makes products safe without regulations	38	39	42	38	40	29	−1	=
Protects jobs of U.S. workers	37	36	41	39	38	27	+1	−2
Working to curb pollution	30	32	31	26	29	21	−2	+4
Solves social problems	22	23	24	21	23	19	−1	+1
Reduces prices when possible	19	19	18	18	18	12	=	+1
Tells all about itself, good and bad	17	13	16	15	16	11	+4	+2

*Multiple responses.

†Adjusted to account for variation in the method of asking the question.

about ITT. A summary of the findings stemming from the surveys conducted during the period 1974 to 1978 is shown in Table 1.

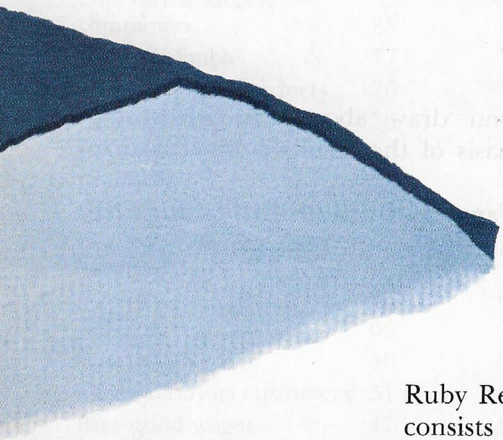
For each attribute listed on the left-hand side of the table, the percentage figures correspond to respondents who felt that ITT possessed the attribute. For instance, while only 68% of the respondents in the January 1974 survey felt that ITT was one of the largest companies, 85% of the respondents in the November 1978 survey felt so.

QUESTIONS

1. What specific inferences can you draw about the advertising campaign's effectiveness on the basis of the results summarized in Table 1? Justify your answer.
2. What additional types of data and analyses might be helpful for evaluating the campaign's effectiveness?

CASE 6.3

Ruby Restaurant



Ruby Restaurant is located in a town in which 25% of the population consists of students attending a major university. The restaurant is owned by Fred Barnes, who graduated with a bachelor's degree in business administration from the university several years ago. It is open only for dinner every day from 5 p.m. until midnight. Well known for its fine food and elegant dining facilities, this restaurant usually attracts patrons considered to be upscale from within and outside town. Despite the above-average prices of Ruby Restaurant's menu items, the restaurant's sales revenue has been growing at a rate considered to be quite satisfactory by Mr. Barnes. Of late, however, he has been somewhat concerned about rumors that two new fashionable restaurants are set to open in the area within the next year or two.

As a first step in designing an effective strategy to offset any potential threat from other new restaurants, Mr. Barnes decided to conduct a survey of his current customers to gain insight into their demographic characteristics and their perceptions about and loyalty toward Ruby Restaurant.

Mr. Barnes designed a brief questionnaire that was handed out to restaurant patrons after they had finished dinner and were waiting for their checks. A total of 878 questionnaires were distributed to all customers who came to Ruby Restaurant during a seven-day period in April. If a party of two or more customers arrived, just one questionnaire was handed out and the party "head" was requested to complete it. As an incentive for participation, the questionnaire offered $2 off the food bill if the customer completed it.

Of the 878 questionnaires distributed, 625 were filled out completely. But on every questionnaire, irrespective of whether or not it was com-

Exhibit 1 *Ruby Restaurant Survey Questions*

1. How far did you travel to reach this restaurant today?

_______ 0 to 3 miles

_______ 4 to 6 miles

_______ 7 to 9 miles

_______ 10 or more miles

2. How many times have you been to this restaurant before today?

_______ I have not been here before

_______ 1 to 3 times

_______ 4 to 6 times

_______ 7 to 9 times

_______ 10 or more times

3. Do you plan to come to this restaurant again within a month?

_______ Yes _______ No _______ Don't know

4. How did you first learn about Ruby Restaurant?

_______ Radio advertisement

_______ Newspaper advertisement

_______ TV commercial

_______ From a friend or relative

_______ From a business associate

_______ Drove by and saw the restaurant

_______ Other
(please specify)

(Continued)

pleted by a customer, Mr. Barnes or one of his assistants recorded the following two pieces of data: number of individuals in the party to which the questionnaire was handed out, and total food bill for the party. These data were recorded on a part of the questionnaire labeled "For Office Use Only." The questions on the survey instrument are shown in Exhibit 1.

QUESTIONS

1. For each question in Exhibit 1 (including question 11), indicate the level of measurement for the collected data (i.e., nominal, ordinal, interval, or ratio).

Exhibit 1 *Continued*

5. Rank the following attributes 1 through 6 in terms of which attribute you feel Ruby Restaurant is best on, which attribute it is second best on, and so on.

_________ Menu variety

_________ Food quality

_________ Restaurant location

_________ Prices

_________ Service

_________ Decor/atmosphere

6. Rank the following attributes 1 through 6 in terms of which attribute is most important to you in selecting a restaurant for dinner, which attribute is next most important, and so on.

_________ Menu variety

_________ Food quality

_________ Restaurant location

_________ Prices

_________ Service

_________ Decor/atmosphere

7. Compared with other restaurants in which you have eaten dinner during the past several months, how would you rate Ruby Restaurant along the following dimensions? (Please circle the appropriate number for each dimension.)

Expensive	1	2	3	4	5	Inexpensive
Friendly	1	2	3	4	5	Unfriendly
Quiet	1	2	3	4	5	Noisy
Good service	1	2	3	4	5	Bad service
Pleasing atmosphere	1	2	3	4	5	Unpleasing atmosphere
Excellent food	1	2	3	4	5	Terrible food

(Continued)

2. State from which questions in Exhibit 1 you would use data, and identify the specific analyses you would conduct on such data, to obtain information pertaining to the following items:

a. Are the customers who filled out the questionnaire different from those who did not?

b. Are customers' impressions of Ruby Restaurant along various attributes asociated with how important those attributes are to them in choosing a restaurant for dinner?

Exhibit 1 *Continued*

8. Your sex is:

______ Male ______ Female

9. Your age is:

______ 22 or under
______ 23 to 30
______ 31 to 45
______ 46 to 60
______ 61 or over

10. The highest educational level you have attained is:

______ Less than high school
______ High school diploma
______ Bachelor's degree
______ Master's degree
______ Doctorate degree

11. For Office Use Only

(a) ______ (b) ______

c. How much customer loyalty does Ruby Restaurant enjoy?
d. What is the most effective medium for letting prospective customers know about Ruby Restaurant?
e. Do customer perceptions of Ruby Restaurant vary in terms of whether the customers are (i) male or female; (ii) under or over 30 years of age; (iii) college graduates or not; (iv) from within or outside town (Ruby Restaurant has a central location and is no more than 6 miles from any part of town)?
f. Do customers' impressions about Ruby Restaurant relative to their impressions about other restaurants have a bearing on how much they spend at Ruby Restaurant?

CASE 6.4

Hamilton Bank (B)

The case titled "Hamilton Bank (A)" described a research proposal submitted by the director of market research to the head of the Commercial Division of the bank. The proposal was approved by the Commercial Division head, and the project got underway a few days later. When all the proposed data had been collected, the director of market research sent a brief interim report titled "Commercial Image Study" to the Commercial Division head. This report is reproduced in full next.

I. Background on our bank.

- Two- (2) billion-dollar bank.
- Headquartered in Lancaster, Pa.
- We have a state charter and deal in a six-county area of 1,350,000 persons.

II. Request.

- Was submitted from the commercial side of the bank.
- The purposes of this study are as follows:
  - a. To determine how our commercial customers perceive us as a supplier of commercial services and also as compared with our major competition.
  - b. To measure the media habits of individuals who make the banking decisions in companies.
  - c. To find out what services are currently being used and what services companies are looking to use in the future.

The information in this case was provided by courtesy of Hamilton Bank, P.O. Box 3959, Lancaster, PA 17604.

  d. To measure brand, competitive, and total image of the commercial side of Hamilton Bank.

III. Methodology logic.

- We used an outside consultant to help us with the study for two reasons:
  a. He is a graduate student working on his Master's Degree and was able to use this project in the writing of his thesis.
  b. The issues discussed were sensitive, and we felt that by using a consultant we could increase the return rate.
- A list of SICs and company names was purchased from an outside supplier. This list was used in the pulling of our sample.
- We randomly selected 1750 businesses from the list: 250 from each of seven (7) counties (6 counties we currently do business in; the 7th we are considering dealing in.)
- Both Hamilton Bank customers and noncustomers were questioned.
- A mail survey was chosen because of the questionnaire's length and of appearance reasons.
- The consultant pretested the questionnaire by conducting ten personal interviews.
- Because of the nature of the study, we had to contact the company before mailing the questionnaire in order to obtain the name of the individual who made the banking decisions for the firm, since this person was the one we wanted to complete the questionnaire.
- The questionnaire was mailed to the firm, along with a cover letter (which was hand-signed).
- Each survey was coded by an 8-digit number: the first 4 digits represented the SIC and the second 4 digits were used as a counter.
- We used stamps for mailing—not bulk mail—and we included a self-addressed, stamped return envelope.

IV. Things about the questionnaire.

- Divided into three sections:
  a. First section—to get a feeling for the size of the company and its media habits.
  b. Second section—to measure image of both the bank and the services in relation to the competition.
  c. Third section—to measure rates of use of specific commercial services.
- The banks and media names change with each county.

V. Uses of data.

- We have a voluminous amount of data that has not yet been analyzed.
- Three tests are being used to analyze the data:
  a. Cross-tabs of SIC with questions in the first section.
  b. A *t*-test of significance for the second section.
  c. Discriminate tests for the third section.
- Each section is being examined by county and in total.
- We received a total response rate of 60% and a usable return rate of 30%. We believe this return rate had something to do with using an outside consultant.
- The result of this study will be a measure of brand, competitive, and total image.
- The information obtained from this study will be used for the following purposes:
  a. To direct calling officers within each county; to answer questions such as, "What do business people feel is important as far as services go?"
  b. To uncover areas of services that need to be developed to meet the needs of businesses in the market; that is, where are we going in the market, and where should we be going? More specifically, it will show us our strengths and weaknesses.
  c. To aid us in developing our advertising attack—that is, where, how, and what to advertise.
  d. To aid us in developing and maintaining a total marketing plan.

QUESTIONS

1. Discuss the strengths—and weaknesses, if any—of the methodology used to gather the data.
2. Critically evaluate the points mentioned in the "Uses of data" section. Which points are clear and which are not? What is your overall assessment of the proposed data analysis implied in this section?

CASE 6.5

National Insurance Company

Tom Kurtis, Manager of Customer Service for National Insurance Company, and his assistant, Jill Baxter, were getting ready to analyze data collected through a mail survey of a sample of National's customers. The primary purpose of the survey was to ascertain how customers perceived National's quality of service and to identify areas for improvement.

THE COMPANY

National Insurance Company is one of the leading insurance firms in the United States. It serves over 10 million customers who have insurance coverage from National for a variety of needs (e.g., home, health, auto, life). The company's customer base, after growing steadily for a number of years, had started to erode during the past two years. Jack Driver, president of National, was concerned about this downturn. Suspecting that there might be some shortfalls in the service National was providing its customers, Jack had asked Tom to conduct a study to determine the "image" that customers had about National's service.

THE SURVEY

In designing the content of the survey, Tom Kurtis drew upon insights he had gained by reading a number of recent articles emphasizing the importance of service quality and suggesting approaches for measuring customers' perceptions of a company's service. In particular, one journal article that was based on extensive customer research suggested that there were five broad dimensions along which customers evaluated service quality.

1. *Tangibles*: The appearance of physical facilities, equipment, personnel, and communications materials.
2. *Reliability*. Ability to perform the promised service dependably and accurately.
3. *Responsiveness:* Willingness to help customers and provide prompt service.
4. *Assurance:* Knowledge and courtesy of employees and their ability to inspire trust and confidence.
5. *Empathy:* Caring, individualized attention the firm provides its customers.

The article also contained a number of illustrative questions pertaining to each of these dimensions and their relative importance. Tom Kurtis and Jill Baxter discussed the contents of this article vis-à-vis their survey objectives and concuded that the suggestions in the article could be adapted for their study. Using the ideas in the article, they designed the questionnaire shown in Exhibit 1.

The items in Section 1 of the questionnaire pertain to customers' evaluation of National's service along the five dimensions. Section 2 is designed to determine the relative importance of the five dimensions. Section 3 contains several questions relating to customers' overall experience with National. The final section focuses on respondents' demographic characteristics.

THE SAMPLE

The questionnaire was mailed to a random sample of 1000 policyholders from National's customer base. On the basis of data available in the company's records on policyholders, the current customer base had the following profile:

Number of years with the company	
Less than 1 year	8%
1 to less than 2 years	9%
2 to less than 5 years	18%
5 years or more	65%
Gender	
Male	54%
Female	46%
Age	
Under 25	5%
25–44	70%
45–64	15%
65 and over	10%

Exhibit 1 *Survey Questionnaire*

Note: The labels for the five dimensions in Sections 1 and 2 did not appear in the original questionnaire. They are included here to indicate the dimension to which each question pertains.

SECTION 1

Directions: The following set of statements relates to your feelings about National Insurance Company. For each statement, please show the extent to which you believe National has the feature described by the statement. Circling a "1" means that you strongly disagree that National has that feature, and circling a "7" means that you strongly agree. You may circle any of the numbers in between that show how strong your feelings are. There are no right or wrong answers—all we are interested in is a number that best shows your perceptions about National.

Reliability	**Strongly Disagree**						**Strongly Agree**
1. When National promises to do something, it does so.	1	2	3	4	5	6	7
2. When you have a problem, National shows a sincere interest in solving it.	1	2	3	4	5	6	7
3. National performs the service right the first time.	1	2	3	4	5	6	7
4. National provides its services at the time it promises to do so.	1	2	3	4	5	6	7
5. National maintains error-free records.	1	2	3	4	5	6	7
Empathy							
6. National treats you with care.	1	2	3	4	5	6	7
7. National has operating hours convenient to all its policyholders.	1	2	3	4	5	6	7
8. National has employees who give you personal attention.	1	2	3	4	5	6	7
9. National has your best interests in mind.	1	2	3	4	5	6	7
10. Employees of National understand your specific needs.	1	2	3	4	5	6	7

(Continued)

Exhibit 1 *Continued*

		Strongly Disagree						Strongly Agree
Reliability								
Tangibles								
11.	National has modern-looking equipment.	1	2	3	4	5	6	7
12.	National's physical facilities are visually appealing.	1	2	3	4	5	6	7
13.	National's employees are neat-appearing.	1	2	3	4	5	6	7
14.	Materials associated with service (such as pamphlets or statements) are visually appealing at National.	1	2	3	4	5	6	7
Responsiveness								
15.	Employees of National tell you exactly when services will be performed.	1	2	3	4	5	6	7
16.	Employees of National give you prompt service.	1	2	3	4	5	6	7
17.	Employees of National are always willing to help you.	1	2	3	4	5	6	7
18.	Employees of National are never too busy to respond to your requests.	1	2	3	4	5	6	7
Assurance								
19.	The behavior of employees of National instills confidence in you.	1	2	3	4	5	6	7
20.	You feel safe in your transactions with National.	1	2	3	4	5	6	7
21.	Employees of National are consistently courteous with you.	1	2	3	4	5	6	7
22.	Employees of National have the knowledge to answer your questions.	1	2	3	4	5	6	7

(Continued)

Exhibit 1 *Continued*

SECTION 2

Directions: Listed below are five features pertaining to insurance companies and the services they offer. We would like to know how important each of these features is to you when you evaluate an insurance company's quality of service. Please allocate a total of 100 points among the five features according to your impression about how important each feature is to you—the more important a feature is to you, the more points you should allocate to it. Please ensure that the points you allocate to the five features add up to 100.

Tangibles

1. The appearance of the insurance company's physical facilities, equipment, personnel, and communications materials. _________ points

Reliability

2. The insurance company's ability to perform the promised service dependably and accurately. _________ points

Responsiveness

3. The insurance company's willingness to help policyholders and provide prompt service. _________ points

Assurance

4. The knowledge and courtesy of the insurance company's employees and their ability to convey trust and confidence. _________ points

Empathy

5. The caring, individualized attention the insurance company provides to its policyholders. _________ points

Total points allocated **100 points**

SECTION 3

For questions in this section, please circle the number that corresponds to your answer.

1. On a scale of 1 to 10, how would you rate the overall quality of service provided by National?

Extremely Poor									Extremely Good
1	2	3	4	5	6	7	8	9	10

2. Would you recommend National to a friend interested in insurance services?

Yes	1
No	2

(Continued)

Exhibit 1 *Continued*

SECTION 3

3. How long have you been using the services of National?

Less than 1 year	1
1 to less than 2 years	2
2 to less than 5 years	3
5 years or more	4

4. Have you recently had a problem with the service you received from National?

Yes	1
No	2

5. If you did have a problem, was it resolved to your satisfaction?

Yes	1
No	2

FINAL SECTION

The following information is solely for statistical purposes and will be kept confidential. Please circle the number that corresponds to your answer.

1. Your sex is:

Male	1
Female	2

2. Your marital status is:

Single	1
Married	2
Widowed	3
Divorced	4

3. Your age is:

Under 25	1
25–44	2
45–64	3
65 or over	4

4. Your total annual family income is:

Under $10,000	1
$10,000–$19,999	2
$20,000–$29,999	3
$30,000–$49,999	4
$50,000–$64,999	5
$65,000 or over	6

5. The highest level of schooling you have completed is:

High school or less	1
Some college	2
College graduate	3
Graduate school	4

Thank you very much!

Completed questionnaires were received from 285 respondents. Being aware that mail surveys typically have low response rates, Tom and Jill were quite satisfied with the response rate of 28.5%. However, they wanted to determine whether the sample of respondents differed in any significant way from National's customer base so that appropriate precautions could be taken in interpreting the findings from the survey.

Assume that you have been asked to assist Tom and Jill in analyzing and interpreting the survey data. In addition to addressing the specific questions listed below, design and conduct other analyses that you think might generate insightful information. To carry out the various analyses, you can obtain from your instructor a copy of the data set and the coding scheme used in generating it.

QUESTIONS

1. Develop and briefly discuss the demographic profile of the survey respondents.
2. Do the survey respondents differ significantly from National's customer base? If there are differences, what bearing do they have on interpreting the findings about customers' perceptions of National's service quality?
3. What is National's average service-perception rating along each of the five dimensions?
4. Which are the most and least critical service dimensions as perceived by customers?
5. What are the correlations between respondents' overall service-quality perceptions (on the 10-point scale) and their average ratings along each of the five dimensions? Are these correlations statistically significant?
6. Conduct a multiple regression analysis wherein overall service-quality perception is the dependent variable and the average ratings along the five dimensions are the independent variables. What specific inferences can you make from the results of this analysis?
7. Are there any significant differences between males and females in terms of how they perceive National's service quality?
8. Is there a significant relationship between customers' willingness to recommend National to a friend and the following?
   a. Their overall service-quality perception
   b. Any of their demographic characteristics
9. On the basis of your answers to the previous questions and any additional analyses you may have conducted, what recommendations would you make to National's managers?

CASE 6.6

Fantastic Car Wash

Fantastic Car Wash is a full-service car wash and detail center in a college town of about 120,000 residents. The only other major car wash and detail center in the area is located about six miles from Fantastic Car Wash. Eric Johnson, owner and manager, founded the company four years ago with an initial investment of $1 million consisting of $200,000 in personal funds and $800,000 in borrowed capital. After barely breaking even in the first two years of operation, the car wash made a net profit of $60,000 in the third year. However, it incurred a loss of $15,000 in the fourth year. Mr. Johnson noticed that the loss was primarily a result of a 20% drop in the number of cars serviced by Fantastic Car Wash during the past year. Alarmed by this sharp decline in business, he hired a small local consulting firm, Dalton Consulting, Inc., to analyze the following issues and make recommendations for improving Fantastic Car Wash's performance:

a. The nature of the car wash industry and recent trends in it.
b. The demographic and other relevant characteristics of Fantastic Car Wash's customers.
c. Customers' perceptions and evaluation of the services provided by Fantastic Car Wash.
d. The effectiveness of Fantastic Car Wash's current promotional expenditures of $25,000 per year, which was approximately 3% of sales. These expenditures were split between television (50%), radio (30%), and discount coupons (20%). All television advertising was on a local channel affiliated with CBS. Radio spots were divided equally among what Eric considered to be the three major stations in the area; KICK (country and western), KROC (top 40), and KSAN (easy listening). The discount coupons were distributed primarily through coupon books that contained similar offers from a number of area retailers.

THE CAR WASH INDUSTRY

After examining a variety of publications and articles pertaining to the car wash industry, Dalton Consulting uncovered several key insights about the industry: Competition in the industry was intense and involved four general groups of competitors—full-service car washes, drive-through rollover car washes, coin-operated spray washes, and do-it-yourself car washes. There were an estimated 23,000 car wash facilities in the United States. The number of facilities had not changed significantly in the past 20 years, with the number of car washes going out of business (about 16% per year) approximately equaling the number of new car washes. A significant trend in the industry was a major shift in the type of car wash used by customers. Large gas stations were installing rollover units in many retail outlets and offering free or discounted washes with the purchase of gasoline. Therefore full-service outlets such as Fantastic Car Wash were under increasing competitive pressure.

Full-service car washes also faced several uncontrollable factors that could significantly affect their profitability. One primary factor was the weather. Car washes typically shut down when it was raining or snowing. Yet, they were expected to be fully operational as soon as the weather improved. Such variability could adversely affect their labor productivity, cash flow, and profits. Another critical factor affecting the car wash industry was the highly seasonal nature of its business. Peak times for this industry were between November and March, when temperatures were low. Warm weather brought increasing competition from do-it-yourselfers.

CUSTOMER SURVEY

To address the company-specific issues raised by Mr. Johnson (i.e., issues b, c, and d listed earlier), Dalton Consulting decided to survey a sample of Fantastic Car Wash's customers. A two-page questionnaire (shown in Exhibit 1) was designed. The actual questionnaire was printed on two sides of a single sheet of paper, with questions 1 through 9e appearing on the front side of the questionnaire and questions 9f through 16 appearing on the back side.

Customers had to wait at least 10 minutes before their cars would be ready. Therefore, Dalton Consulting asked the cashier, who was stationed in the customer waiting area, to distribute the questionnaire to customers to be filled out by them while they waited. Specifically, the cashier was instructed to hand out a questionnaire to every tenth customer during the first two weeks of April. A total of 280 questionnaires were handed out over a period of 14 days during operating hours. Each customer receiving a questionnaire was requested to complete the questionnaire and drop it off in a box that was kept next to the cashier's window. At the end of the survey period the box contained a total of

Exhibit 1 *Survey Questionnaire*

To help us serve you better, please take a few minutes to fill out this short questionnaire. Thank you for your help.

1. How often do you usually have your car washed?

_______ Less than once a month
_______ Once a month
_______ Two times a month
_______ Three times a month
_______ Four or more times a month

2. How often do you usually have your car waxed?

_______ Less than once a month
_______ Once a month
_______ Two times a month
_______ Three times a month
_______ Four or more times a month

3. How often do you usually wash and/or wax your car by yourself?

_______ Less than once a month
_______ Once a month
_______ Two times a month
_______ Three times a month
_______ Four or more times a month

4. How long has it been since you last brought your car to Fantastic Car Wash?

_______ Less than a week
_______ One to less than two weeks
_______ Two to less than three weeks
_______ Three to less than four weeks
_______ Four weeks or more

5. How far away from Fantastic Car Wash do you live?

_______ Less than one mile
_______ One to less than three miles
_______ Three to less than five miles
_______ Five miles or more

(Continued)

Exhibit 1 *Continued*

6. How old is your car that is now being washed/waxed at Fantastic Car Wash?

________ Less than one year

________ One to less than two years

________ Two to less than three years

________ Three to less than four years

________ Four years or more

7. In the past three months, have you used any car wash other than Fantastic Car Wash?

________ Yes ________ No

8. How long are you willing to wait to have your car professionally washed/waxed? ________ minutes

9. Indicate the extent to which you disagree or agree with the following statements. For each statement, circle "1" if you strongly disagree, circle "7" if you strongly agree, or circle an appropriate number in between if your feelings are less strong.

	Strongly Disagree						**Strongly Agree**
a. The washing/waxing job done by Fantastic Car Wash is excellent.	1	2	3	4	5	6	7
b. The service performance of Fantastic Car Wash is consistent from one visit to the next.	1	2	3	4	5	6	7
c. The prices charged for Fantastic Car Wash's services are reasonable.	1	2	3	4	5	6	7
d. The time taken by Fantastic Car Wash to provide its services is reasonable.	1	2	3	4	5	6	7
e. The customer waiting area at Fantastic Car Wash is comfortable.	1	2	3	4	5	6	7
f. I would recommend Fantastic Car Wash to anyone looking for a professional car wash.	1	2	3	4	5	6	7
g. I would rate Fantastic Car Wash as being excellent overall.	1	2	3	4	5	6	7

(Continued)

Exhibit 1 *Continued*

---

10. What is you age?

__________ 16–20

__________ 21–25

__________ 26–30

__________ 31–40

__________ 41–50

__________ 51 or over

11. What is your sex?

__________ Male __________ Female

12. What is your primary occupation? ____________________

13. What is your marital status?

__________ Married __________ Single (Please skip to question 15)

14. If married, what is your annual household income?

__________ Under $10,000

__________ $10,000–$19,999

__________ $20,000–$29,999

__________ $30,000 or more (Please skip to question 16)

15. If single, what is your annual personal income?

__________ Under $10,000

__________ $10,000–$19,999

__________ $20,000–$29,999

__________ $30,000 or more

16. What newspaper do you read *most* often? ____________________

17. What radio station do you listen to *most* often? ____________________

18. What TV station do you watch *most* often? ____________________

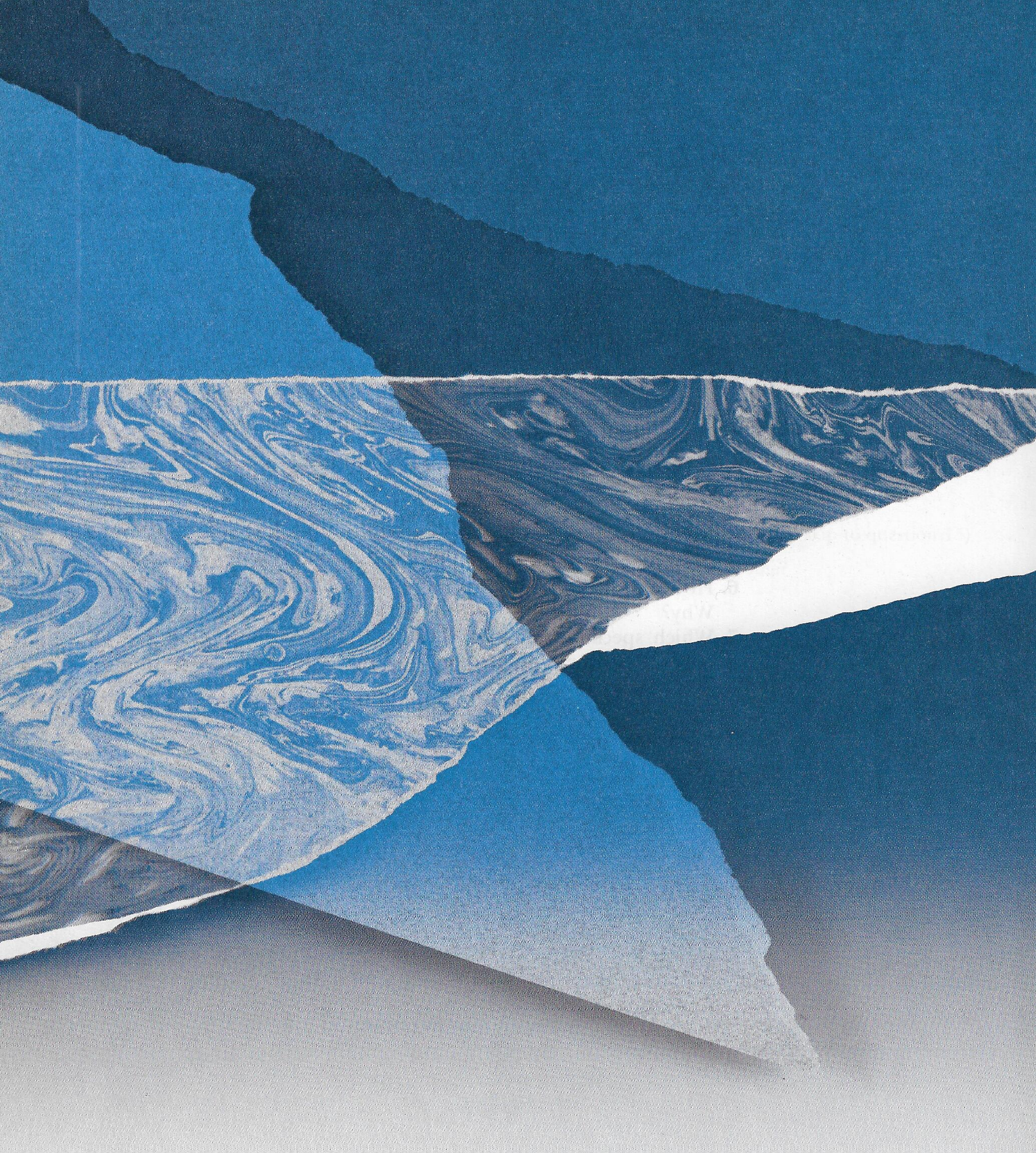
Thank you very much!

256 completed questionnaires. The remaining 24 questionnaires were apparently discarded by the customers who received them.

Dalton Consulting had coded the data from the 256 questionnaires and was getting ready to analyze them so as to address the specific issues raised by Mr. Johnson. (A copy of this data set and the coding scheme used is available from your instructor.)

QUESTIONS

1. Critically evaluate the questionnaire designed by Dalton Consulting. What changes, if any, would you recommend and why?
2. Is the sampling approach used to administer the survey appropriate? Discuss the reasons for your answer.
3. What analyses would you conduct using the collected data in order to generate useful information for Mr. Johnson?
4. How would you describe the profile of the customer base that was surveyed?
5. On the basis of customer perceptions, what are Fantastic Car Wash's areas of strength? What are its areas of weakness?
6. How effective is Fantastic Car Wash's current promotional strategy? Why?
7. Which specific attributes of Fantastic Car Wash seem to influence customers' overall evaluation of the firm and their willingness to recommend the firm?
8. Based on the results of your various analyses, what recommendations would you make to Mr. Johnson *and* what caveats would you include with them?



PART SEVEN

COMMUNICATING with Research Users

CHAPTER 20 *Presentation of Research Results*

CHAPTER 20

Presentation of Research Results

CHAPTER OBJECTIVES

After understanding the material in this chapter, you should be able to do the following:

- *Discuss why it is critical to understand the nature of the audience before designing a written report or an oral presentation.*
- *Describe the key elements included under each of the desirable report features summarized in the mnemonic word SIMPLE (short, interesting, methodical, precise, lucid, and error-free).*
- *State the components of a written report and indicate what each component should contain.*
- *Recast statements containing statistical or technical terms into plain language.*
- *Define the following graphical illustrations and indicate when each is likely to be most appropriate: pie chart, line chart, stratum chart, bar chart, pictogram.*
- *Graphically represent information contained in one-way and two-way tabulations.*
- *Identify unique oral-presentation characteristics that call for special attention and effort during preparation.*

Communicating the results of a research project to decision makers is *the last step of the research process. Nevertheless, the position* of this topic in the sequence of research steps is no indication of its importance. Indeed, whether and to what extent research findings have an impact on decision making critically depends on how well they are communicated. As we have seen in earlier chapters, a research project is worthless if decision makers do not act on its findings, no matter how accurately and thoroughly it is conducted.

Effective presentation is the critical difference between a written report's gathering dust—or an oral presentation's falling on deaf ears—and the research findings' being truly beneficial. Unfortunately, communication is perhaps the weakest link between researchers and research users—a point emphasized at the beginning of the book. The topics of this chapter should therefore be studied carefully, especially by budding researchers who may be captivated by the methodological aspects of marketing research and thus may neglect the aspect of communicating their findings effectively. When a sample of marketing managers were asked to give their best advice to college students interested in marketing research careers, many of them gave answers like the following:

> Develop a good ability to communicate well—both orally and in writing.
>
> Work very hard on written and verbal communication skills, especially on how to say a great deal in as few a number of words as possible.[1]

---

[1] A. Parasuraman and Deborah G. Wright, "A Study of Marketing Research Jobs for College Graduates: Implications for Educators," in *1983 AMA Educators' Proceedings,* eds. Patrick E. Murphy, Gene R. Laczniak, Paul F. Anderson, Russell W. Belk, O. C. Ferrell, Robert F. Lusch, Terence A. Shimp, and Charles B. Weinberg (Chicago: American Marketing Association, 1983), p. 184.

IMPORTANCE OF UNDERSTANDING THE AUDIENCE

Researchers invariably present a project's description and findings in the form of a written report. Frequently, they also supplement the written report with an oral presentation to highlight key findings and answer any questions decision makers may have. Regardless of whether a research is summarized in written or oral form, however, its potential impact depends on how well it is tailored to fit the background and needs of its audience. A report or presentation that is prepared without considering its intended audience is bound to be ineffective. While this point appears to be obvious, in real life it is not always heeded. As we saw in Chapter 1, researchers sometimes attempt to impress rather than *communicate with* decision makers. Misunderstanding between research providers and users is the inevitable result.

Gaining a good understanding of audience characteristics and needs is therefore the foremost step in writing a research report or preparing an oral presentation. Thus a researcher should explore questions like these: Individuals from which organizational levels are likely to read the report or attend the oral presentation? How busy are these individuals likely to be? How familiar are they with the project? What aspects of the project are they most likely to be interested in? Do they have the background and training to easily understand technical complexities and terminology related to the project?

Researchers may not always be able to find accurate answers to questions like the ones just posed. Nevertheless, being sensitive to these questions is a prerequisite for ensuring a good report or presentation. Moreover, when the audience consists of multiple groups with widely disparate backgrounds and interests, a researcher should prepare several separate reports or presentations, if feasible, in order to maximize the overall impact of a research project. The importance of being user-oriented at this final and *most critical* stage of a research project will become more evident in our subsequent discussion.[2]

The primary focus of this chapter will be on written reports. However, most of the points covered are germane to oral presentations as well. Certain guidelines critical for making good oral presentations will be stressed at the end of the chapter.

PREPARING EFFECTIVE WRITTEN REPORTS

Writing skills cannot be acquired overnight; they have to be cultivated through constant practice over a period of time. Similarly, the details for preparing good written reports cannot be covered in a single chapter;

---

[2] The importance of understanding the audience and how to do so effectively are discussed succinctly by Dianna Booher, *Would You Put That in Writing?* (New York: Facts on File, 1983), pp. 15–19.

numerous comprehensive books and manuals on report writing are available for that purpose. Therefore, our purpose here is merely to offer a checklist of key features that an effective research report should contain and to illustrate the importance of those features. Details on the variety of ways in which such features can be incorporated into a report are available in several of the references cited in this chapter.

The desirable features of a research report can be summarized in one mnemonic word, SIMPLE: *s*hort, *i*nteresting, *m*ethodical, *p*recise, *l*ucid, and *e*rror-free. These six features do overlap to some degree. However, each is important enough to be discussed separately.

Make It Short

Decision makers usually have no more than a few hours to read a report summarizing a study that may have taken weeks or months to complete. Hence the report must be as succinct as possible. It should cover in detail only those aspects that are of interest to research users and have a direct bearing on their decision making. This task is more easily said than done, however. Having labored for long on a project, researchers may be tempted to explain everything they did, perhaps to obtain the audience's appreciation for their efforts. This temptation is likely to be strong among those unaccustomed to writing research reports. In his excellent book on report writing, William J. Gallagher states:

> In their exuberance, many inexperienced writers find it difficult to screen out irrelevancy. Anxious to display their new-found knowledge and laboring under the delusion that their education is unique, recent graduates of advanced programs in science and business administration are often guilty of including unnecessarily elaborate discussions of theory and technique even though the theory may be well-known and the method conventional. Other writers are often victims of a kind of narcissism. For example, they write at great length about the difficulty they encountered in obtaining the data, the implication being that only their ingenuity and persistence could have avoided abysmal failure. Others write *ad nauseam* about the sample selected even though it is immediately apparent that the sample is representative and valid. Still others include a wealth of information about apparatus and methods to protect themselves against the carping criticism of other members of their profession, even though these people are not included among the intended readers of the report.[3]

This quotation does not necessarily mean that technical or methodological details must be omitted from research reports. All it implies is that report writers must set aside their personal or psychological needs and focus on potential readers' needs instead. A report should, of course, include technical details if its readers are likely to be interested in them.

---

[3] William J. Gallagher, *Report Writing for Management* (Reading, Mass.: Addison-Wesley, 1969), pp. 87–88.

As mentioned earlier, a thorough understanding of the audience is essential for producing a concise, effective report.

How concise should a report be in order to be labeled "short"? There is no standard length for a report. The ideal length depends on the needs of the audience and hence will vary from one situation to another. The intent of the discussion in this section is not to suggest a standard page limit for research reports. Rather, it is to emphasize that a report must not contain material irrelevant to its intended audience—a piece of advice that is ignored more often than it is followed by report writers. In his article on writing good reports, Stewart Henderson Britt, a respected marketing scholar, observes:

> Errors of *too long* a report far exceed the number of instances of *too short* a report. You do not have to report every detail, and neither should you always give every finding equal relevance. You do not even have to have a table or a graph accompanying every statement of results, unless this adds to the information needed by your readers.[4]

Make It Interesting

The editorial philosophy of *Decision Sciences,* a technical journal, offers the following advice to would-be contributors of articles:

> Authors should especially bear in mind the fact that the journal is not intended as a cure for insomnia, so the occasional injection of a little wit and/or thought-provoking asides will not be discouraged.[5]

This advice is relevant to all report writers, particularly to those prone to inundating their reports with confusing statistics, tables, and formulas. While a report does not have to be humorous or entertaining, it must be interesting enough to grab and maintain the attention of its readers. As we saw in Chapter 5, the current information explosion is making many managers drown in data. There is just too much competition for the time that managers can devote to research reports. Therefore an uninteresting report may not even be fully read, let alone have an impact on decision making.

How does one prepare an interesting report? A researcher can prepare an interesting report by understanding the audience and by writing in a style appealing to them;[6] by focusing the report's content on issues that are relevant to the audience and important for it to know, a point we stressed in the previous section; and by using suitable illustrations and visual aids, a topic we will discuss later.

---

[4] Stewart Henderson Britt, "The Writing of Readable Research Reports," *Journal of Marketing Research,* 8 (May 1971), p. 266.

[5] "Editorial Philosophy," *Decision Sciences,* 16 (Spring 1985), p. vi.

[6] In addition to Gallagher's book, *Report Writing for Management,* a number of other books on writing style are available, such as Paul R. Timm, *Managerial Communication: A Finger on the Pulse* (Englewood Cliffs, N.J.: Prentice-Hall, 1980); Jessamon Dawe and William Jackson Lord, Jr., *Functional Business Communication* (Englewood Cliffs, N.J.: Prentice-Hall, 1968); William C. Himstreet and Wayne Murlin Baty, *Business Communications* (Belmont, Calif.: Wadsworth, 1973); Dianna Booher, *How to Write Your Way to Success in Business* (New York: Harper & Row, 1984); Paul V. Anderson, *Business Communication: An Audience-Centered Approach* (Orlando, Fla.: Harcourt Brace Jovanovich, 1989).

Make It Methodical

A written report must be well organized to avoid readers' confusion and facilitate their quick comprehension. The various sections within the report must be logically sequenced to ensure smooth transitions. The number of sections and the best way to arrange them depend to some degree on the nature of the project and of the audience. However, the following general outline should be helpful for structuring reports in most situations.[7]

1. Preliminary sections
   a. Transmittal letter
   b. Title page
   c. Table of contents
   d. Executive summary
2. Body of the report
   a. Purpose
   b. Method
   c. Findings
   d. Conclusions (and recommendations if requested)
   e. *Limitations*
3. Addenda
   a. List of references
   b. Appendix

Not all sections listed in this outline are necessary in every report. The following description of the outline discusses the role of the various sections.

Transmittal Letter. As its name implies, the transmittal letter introduces the report to its audience. It typically follows the style of business letters and includes a brief description of the report's highlights. The contents of the executive summary section, which we will discuss shortly, can be incorporated into the transmittal letter if the entire report is brief. The transmittal letter is optional when the researcher personally delivers the report to its readers.

Title Page. The title page contains the title of the study, the name/affiliation of the report's author, the date of the study, and the name/

---

[7] This outline is adapted from Himstreet and Baty, *Business Communications,* p. 335; many ideas in the descriptions of individual components of the outline are also adapted from this source. This outline is not the only available outline; examples of some different outlines can be found in Timm, *Managerial Communication,* Dawe and Lord, *Functional Business Communication,* and Booher, *How to Write.*

affiliation of the person requesting the study. An attractively laid out title page is important for making a good first impression on the reader.

Table of Contents. When a report contains many sections, a table of contents helps the reader quickly locate any desired section. The contents page customarily does not include the title of the study; it is labeled "Table of Contents" or simply "Contents." When the body of the report contains numerous tables, figures, or other illustrations, a desirable feature is a separate page titled "List of Tables," "List of Figures," or the like, after the contents page.

Executive Summary. The executive summary is a synopsis of the entire report and typically should be no more than a few pages long. Although an executive summary may not be necessary in every report, it should be included in lengthy and/or technically complex reports. It should describe succinctly the primary purpose of the study, the general methodology used, the significant findings, and the key conclusions/recommendations. It is often the only section of a report that busy executives, who may well be the people with final authority to act on the study's findings, will have time to read thoroughly. A carelessly written executive summary is bound to create a poor image of the study and hurt the chances of its findings having any impact. Writing an executive summary is therefore an important task. Viewing this task as one that can be accomplished in a hurry after completing the main report is incorrect. Ideally, the executive summary should be drafted first by carefully distilling from the study the main messages to be conveyed to readers.

The primary purpose of an executive summary is to communicate concisely the key issues the study focused on, the findings uncovered about each issue, and the implications of those findings. Therefore a useful starting point for drafting an executive summary is to outline the main study objectives (i.e., issues about which information was sought through the study) and list under each objective all the findings that relate to it. Such an outline will facilitate pulling together in one place findings from different sections of the study pertaining to the same objective.

To illustrate, suppose a firm marketing food products commissions a study involving a consumer survey in which a key objective is: "To identify distinguishing features, if any, of users of our brand of cake mix vis-à-vis users of competing brands." Findings from several different sections of the survey questionnaire (e.g., sections focusing on brand preference, purchase frequency, usage behavior, and demographics) will be germane to this objective. Organizing these findings under a common objective and examining them collectively will help identify key insights to be included in the executive summary. For instance, examining one-way and two-way tabulations of responses to questions about brand preference, purchase frequency, usage behavior, and demographics may suggest something like the following for inclusion in the executive summary:

> Users of our brand of cake mix, while similar to users of competing brands in several respects, also have a number of distinguishing characteristics. Specifically, users of the various brands do not differ significantly with respect to age, education, and household size. However, the median income of households using our brand is higher than that of households using competing brands ($20,000 versus $15,000). Moreover, the purchase frequency of our brand is lower than for competing brands (6 versus 10 boxes per household per year). Finally, users of our brand seem to be more brand loyal than users of competing brands (7 out of 10 users of our brand said they are likely to buy the same brand in the future while only 5 out of 10 users of competing brands said so). Thus our brand enjoys an upscale, premium image which has been the goal of our advertising campaign during the past two years. However, for our future campaigns we might consider themes that also promote more frequent purchases by households using our brand.

Main findings and implications pertaining to the various research issues or objectives addressed by a study can be transformed into brief sections such as the one illustrated above. These sections, along with several introductory paragraphs briefly describing the study's purpose and methodology, form the building blocks of an effective executive summary. A concisely written executive summary that organizes the key findings by research objectives can also aid the researcher in drafting the full report: It can serve as a skeleton for structuring and fleshing out the body of the report. In fact, the sequencing of topics covered and results presented in the executive summary should be the same as in the body of the report. Readers who want more information about a particular section can then quickly turn to the corresponding section in the main body of the report.

Body of the Report. The body offers readers a comprehensive picture of the study by discussing all phases of it. The outline presented earlier lists five sections under this segment. These sections comprise the various research-project steps—starting with definition of the project's purpose and concluding with interpretation of the project's findings—discussed earlier in the book. As such, they can be divided into suitable subsections if they become lengthy or complex. While their actual contents will vary from one study to another, they should focus on certain general issues outlined next.

Purpose. This section should provide a clear definition of the research problem and list the study's objectives. A brief discussion of the rationale for the objectives and their relevance to the problem at hand is also appropriate for this section.

Method. This section should describe and justify the procedures used in generating and analyzing the data for the study. Illustrative of the types of issues to be addressed in this section are the following:

- What overall research approach or design (exploratory, conclusive, or exploratory followed by conclusive) was used in the study? Why was this design the most appropriate one?
- Did the study rely solely on secondary data or was primary-data collection also involved? What sources of secondary data were used?
- How were primary data, if any, gathered (through observation, questioning, or both)? What constituted the study population and how was the sample selected from it? If the questioning approach was used, what method was used to administer the questionnaire (mail, telephone, or face-to-face interviews)? During which period were the data collected?
- What techniques were used to analyze the collected data? For instance, were the analyses limited to construction of one-way and two-way tables and computation of summary statistics, or did they employ more advanced statistical procedures?

Findings. This section should present and discuss the results from all the analyses conducted. As was recommended for the executive summary, an effective approach for this section is to group the findings by research objective. However, this section should offer much greater detail than the executive summary. For instance, while the executive summary might present only the key insights from a series of tables, the body of the report should present the tables in their entirety and discuss all the insights they reveal. The use of graphical illustrations (to be discussed a little later) is an effective way of presenting the findings in an economical yet interesting fashion.

Conclusions. This section should briefly review the study, summarize the inferences stemming from it, and highlight the evidence supporting each inference. It typically should be quite brief since its purpose is to recapitulate the main points discussed in earlier sections. This section could also include a set of recommendations if they were requested by whoever commissioned the study. While some managers might want the researcher to make specific recommendations, others might prefer that the researcher simply report and interpret the findings, without recommending how the managers should act on those findings.

Limitations. This section should include any cautionary notes that qualify the results of the study. Any unforeseen or unavoidable methodological weakness or external events that might have a bearing on the study's conclusions should be covered here. For instance, if the demographic composition of the study sample is markedly different from that of the population of interest, this discrepancy must be explicitly acknowledged. Moreover, a brief discussion of why the discrepancy arose and how it might influence the generalizability of the study's findings should be included.

In summary, the body of the report, consisting of the sections described above, is essentially a report within a report. As such, it should be written in accordance with the guidelines implied by the mnemonic SIMPLE.

Addenda. Addenda are relevant attachments referred to, but not included in, the body of a report. A typical addendum is a list of references. This list is necessary when numerous citations—data sources, past studies, articles, books, and so on—are made in the main report. When only a few citations are involved, they can simply be referenced in footnotes on the pages where they appear, thereby eliminating the need for a separate list of references. Another addendum to research reports is an appendix (or set of appendixes). Questionnaires, raw data tables, formulas, complex calculations, and similar items are usually placed at the end of a report as a set of appendixes. A questionnaire containing a summary of responses to each question is a particularly useful appendix because it simultaneously presents the actual questions asked and a bird's-eye view of the answers. Research in Use 20.1 describes illustrative excerpts from such a questionnaire. In general, any study-related material whose placement in the body of a report may be distracting rather than enlightening must be relegated to the appendix section.

Make It Precise

A precise report is a *clear* report, one that is composed of unambiguous statements. It is also a *comprehensive* report, one that contains all the information necessary for readers to get a complete and true picture of the study. The first feature (i.e., clarity) is a function of writing style and report format, which we will discuss further in the following sections.

The second feature (i.e., comprehensiveness) involves providing enough details about the study—for example, nature of the sample, data collection procedures, analysis techniques—to help readers decide how valid and generalizable the study's findings are. Withholding such details can easily mislead readers. In terms of ensuring objectivity as well as being ethical, researchers have an obligation to inform research users about potential limitations of a study.

As we have seen throughout this book, to expect practical research projects to be *perfect* is unrealistic. Experienced research users are well aware of this shortcoming. Consequently, they are bound to view with skepticism reports that sound too good to be true. Contrary to what some researchers may believe, a report that acknowledges a study's lim-

Research in Use 20.1 *Study of Purchasing Professionals*

Questionnaire with Summary of Study Findings

An economical way to present all the data collected through a survey is to include in the research report a copy of the questionnaire containing summary statistics for the various questions (e.g., frequency distributions for questions with categorical responses and means and standard deviations for questions with scaled responses). Shown here are two excerpts from a questionnaire that was included in a report on a survey of purchasing officers conducted by the author. In both excerpts N is the sample size and the other numbers are percentages unless stated otherwise. The following excerpt illustrates the presentation of summary statistics for questions involving rating scales.

To what extent are the *purchasing decisions* that you make, regarding the choice of vendors and products, influenced by each of the following? Please indicate your answer by circling the number that best describes your opinion about each item.

The numbers 1–5 may be described as follows:

1 No extent
2 Limited extent
3 Moderate extent
4 Considerable extent
5 Very large extent

$N = 261$
Mean scores (underlined)

	No Extent	Limited Extent	Moderate Extent	Considerable Extent	Very Large Extent	Std. Dev.
Sales calls made by suppliers' salesmen	1	2	3.10	4	5	1.09
Suppliers' advertisements in trade journals/ magazines	1	2.10	3	4	5	.94
Promotional materials mailed to you by suppliers	1	2.20	3	4	5	.90
Trade shows where suppliers' products are displayed	1	2.47	3	4	5	1.18
Recommendations made by your colleagues in other organizations	1	2	3.04	4	5	1.09
Specifications laid down in the purchase requisitions	1	2	3.84	4	5	1.17

(Continued)

Research in Use 20.1 *Continued*

The next excerpt illustrates the presentation of summary statistics for questions involving categorical or open-ended responses.

Please provide the following background information about yourself.

1. Are you male or female? 89 Male 11 Female N = 260

2. From the following four categories please check the *highest level* of education that you have attained.

42 High school diploma N = 257
48 Bachelor's degree
9 Master's degree
1 Doctorate degree

3. How old were you on your last birthday? ________ years old
Mean = 44; range = 21 to 85; N = 259.

4. What is the length of your experience in the purchasing profession?
________ years and/or ________ months
Mean = 15 years; range = 1 to 68 years; N = 258.

5. Please indicate your approximate yearly salary before taxes: $________
Mean = $31,900; range = 7800 to 200,000; N = 232.

6. Under which of the following categories does your current job title fit? (Please check the *most appropriate* category.)

32 VP or director of purchases/materials management N = 263
12 General purchasing agent
14 Materials manager or manager of purchasing
18 Purchasing agent
4 Supervisory, senior or administrative buyer
3 Buyer
0 Junior or assistant buyer
17 Other ________________________________
(please specify)

↘

President	5%
Store mgr.	3
Owner	2
Treasurer	2
All others	5
	17%

itations is likely to create a better impression about the study than a report that fails to do so.

Report writers must be careful, though, not to overdo the reporting of limitations by listing every little problem the study may have encountered. Doing so will not only lengthen the report unduly but also give the study a poor image. Furthermore, merely acknowledging all the limitations of a study does not make the study a good one. A study that is badly flawed because of researcher carelessness cannot be salvaged simply by detailing all its limitations.

In short, a report summarizing a *reasonably well conducted study*—one whose limitations were beyond the researcher's control—will be more credible when it mentions the key limitations than when it does not.

Make It Lucid

The lucidity of a report is critical to its effectiveness for several reasons. First, the clearer a report, the less time readers need to understand what it says, and the more time they can spend *digesting* its contents. As was mentioned earlier, most managers can devote only a limited amount of time to research reports. Hence the more time they can devote to digesting, rather than simply understanding, a report's contents, the greater the potential impact of the report will be. Second, a report that lacks clarity runs the risk of confusing its readers and leading them to conclusions unwarranted by the study and unintended by the report writer. Third, an unclear report can annoy its readers and create a poor image about the quality of the entire research project.

How does one prepare a clear report? In a general sense, clarity is simply a function of the other desirable report features we have already discussed or will discuss presently. A report that is short, interesting, methodical, precise, and error-free is bound to be clearer than one that does not possess these charcteristics. There are, however, a few other, more specific guidelines for enhancing report clarity.

One prerequisite for developing a clear report is to thoroughly understand the audience and to write in a style compatible with its background. A frequent complaint of managers is that research reports are too technical. The source of this complaint is invariably the report writer's failure to pay adequate attention to the audience. Technically complex material may be clear to the report writer and may even be viewed as a means for impressing the reader. However, more often than not, such material will neither enlighten nor impress the reader. On the contrary, it is most likely to result in dismissal of the report as being too theoretical to be of practical value.

Devoting a little extra time and effort to translating technical statements into plain language can greatly improve report clarity, as the following illustrations demonstrate:

- Technical statement. Owing to severe multicollinearity between the independent variables (personal selling expenditures and advertising expenditures), it is difficult to interpret their regression coefficients (estimated beta values) and hence their impact on the dependent variable (sales).

- Plain-language translation. Since personal selling and advertising expenditures for our product have moved together closely in the past, to say which of these two types of expenditures may be responsible for changes in our product's sales is difficult.
- Technical statement. According to our analysis of the survey data, the difference between mean sales per store in regions A and B is statistically significant beyond an alpha level of .05, based on a two-tailed *t*-test.
- Plain-language translation. According to our analysis of the survey data, the odds are less than 5 in 100 that the observed difference between mean sales per store in regions A and B could have occurred merely by chance.

The preceding examples are not meant to suggest that a report should be written in a style that talks down to its reader. To use technical terminology (especially when it will help reduce a report's length) is appropriate *if* the audience consists of those who are familiar with such terminology. Indeed, explanations of technical aspects in lay terms may run the risk of insulting a sophisticated audience. The bottom line is that the nature of the audience determines how technical a report can get without losing clarity.

A rather straightforward way to improve the clarity of a report, regardless of the technical complexity level of the project it pertains to, is to use simple words and sentence structures.[8] However, the skill of writing in a simple yet effective style can be developed and strengthened only through practice. There are no shortcuts or formulas (although several have been suggested in books on effective written communications) that can guarantee simple and effective reports.[9] As Gallagher points out:

> Many respected writing authorities have inadvertently contributed to the myth of the magic formula by developing statistical norms and mathematical yardsticks based on word, sentence, and paragraph length to determine ease of reading and, by implication, effectiveness of writing. Although these formulas are intended only as broad guidelines, not as stimulants to creative effort, many writers have misinterpreted the intent. As a result, they have divorced content from form, ignored relationships between ideas, and blindly assumed that writing is a mechnical skill that

---

[8] The references cited in footnote 7 offer numerous suggestions for writing reports in a style that is simple and readable.

[9] One formula that is frequently used to assess the clarity of written material is called the Fog Index, developed by Robert Gunning. Details on this formula are given in Robert Gunning, *The Technique of Clear Writing* (New York: McGraw-Hill, 1952).

can be acquired merely by keeping words and sentences short. Therefore, they often become concerned more with words than with thoughts.[10]

Last, but not least, the use of appropriate tables, charts, or figures can add to the clarity of a report by conveying the contents of written material much more effectively and parsimoniously. For instance, consider the following excerpt from an article published in a leading marketing journal (in this excerpt ES1 and ES2 refer to samples selected in two different studies):

> A comparison of the demographic profiles between ES1 and ES2 revealed no major differences. . . . In ES1, the typical respondent was white (85.5%), married (72.7%), male (66.8%), and employed in a white collar job (36.4%). Fifty-five percent of the respondents reported incomes of less than $20,000, 48.2% had completed high school or less, and 44% were under 45 years of age. In ES2, 59.7% were male, 70.2% were married, 28.9% held white collar jobs, and 81.8% were white. Approximately 43% of the sample were under 45, 53.2% reported having a high school education or less, and almost 50% stated that they had family incomes of less than $20,000 a year.[11]

The information conveyed by this excerpt would have been much clearer had it been presented in tabular form, as follows:

Respondent Characteristic	**Percentage of Respondents**	
	ES1 Sample	**ES2 Sample**
White	85.5	81.8
Married	72.7	70.2
Male	66.8	59.7
White-collar employee	36.4	28.9
Income less than $20,000	55.0	50.0
High school education or less	48.2	53.2
Under 45 years of age	44.0	43.0

We will further discuss the use of tables and charts in the section on graphical illustrations.

Make it Error-Free

To make a good impression on readers, a report must be free of arithmetical, grammatical, and typographical errors, however minor they may be. While this point seems obvious, report writers frequently neglect to edit and proofread their work carefully. Checking a completed report thoroughly should take no more than a small fraction of the time taken to put the report together. Yet perhaps because of a premature feeling of relief that the task is done, writers often look over their final drafts hastily or not at all.

---

[10] Gallagher, *Report Writing for Management*, p. 13.

[11] Richard M. Durand, Hugh J. Guffey, Jr., and John M. Planchon, "An Examination of the Random Versus Nonrandom Nature of Item Omissions," *Journal of Marketing Research*, 20 (August 1983), p. 307

The price of carelessness can be quite high, however. Incorrectly typed numbers, percentages that do not add up right, misspelled words, and other minor errors are usually *major* annoyances to readers. These errors will distract readers and divert their attention from digesting a report's contents. Morever, a report with several such errors will make readers skeptical about the quality of the project itself. Checking a report more than once, preferably with the help of someone good at editing and proofreading, is a wise and worthwhile investment of the report writer's time.

In summary, the key to maximizing a report's potential effectiveness is to make it SIMPLE: short, interesting, methodical, precise, lucid, and error-free. The next section describes a variety of ways of presenting data graphically so as to highlight the key insights embedded in them.

GRAPHICAL ILLUSTRATIONS

Research reports typically contain a variety of data whose main messages must be communicated to the audience clearly and quickly. As the example concerning the ES1 and ES2 samples showed, presenting data in tabular form is usually more effective than describing them in the body of a report. A report's clarity can be further improved by recasting the data in tables into graphical illustrations (charts, graphs, etc.), especially when the tables are complex or confusing. Visually appealing graphical illustrations can also add to a report's communication effectiveness by giving its audience refreshing breaks from the monotony of words and numbers.

One's imagination is the only limit on the number of ways tabular data can be transformed into appropriate graphical illustrations. Nevertheless, there are really only a handful of basic charts and graphs, variations of which are frequently used. To describe these basic formats, let us consider the hypothetical data in Table 20.1 pertaining to market shares of six brands of toothpaste in a certain region.

Suppose Table 20.1 is presented to the manager in charge of marketing brand X. Close examination of the table should offer the manager several key insights. For example, the two dominant competitors of brand X are brands A and B, and brand X's market position has been weakening over the years. These insights, and others not immediately apparent from the table, can be communicated more efficiently and effectively through graphical illustrations, such as the ones discussed next.

Table 20.1 *Market Shares of Six Toothpaste Brands*

	Market Shares (%)					
Year	**Brand X**	**Brand A**	**Brand B**	**Brand C**	**Brand D**	**Brand E**
1980	30	29	25	5	5	6
1981	22	35	22	7	6	8
1982	25	25	30	10	5	5
1983	25	29	28	8	6	4
1984	28	30	25	6	6	5
1985	25	28	28	6	7	6
1986	22	30	29	6	6	7
1987	22	27	32	7	7	5
1988	20	29	33	5	7	6
1989	18	28	34	7	8	5

Pie Charts

Pie Chart *A circle divided into several slices whose areas are proportionate to the quantities to be represented on the chart.*

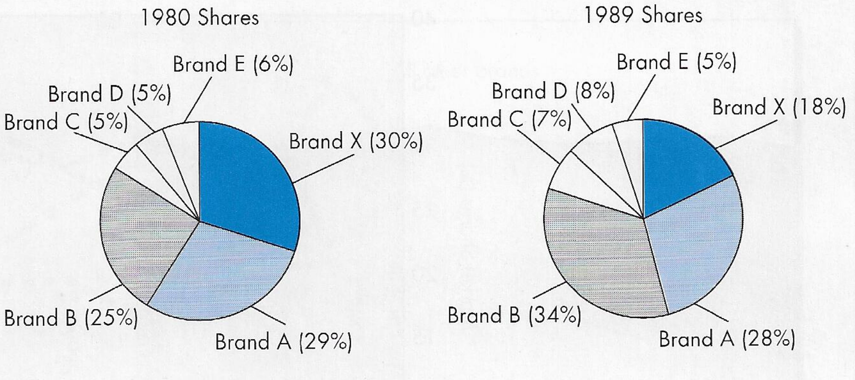
A ***pie chart*** is a circle divided into several slices whose areas are proportionate to the quantities to be represented on the chart. The relative sizes of the slices are shown as percentages. A pie chart is excellent for showing the decomposition of a total quantity into its components—for example, the contributions made by a firm's product lines to its sales or profits; the income distribution of households using a certain product; the allocation of a total budget to various expense categories; the market shares of firms within an industry or brands within a product category. Exhibit 20.1 illustrates two pie charts showing the market shares of the six toothpaste brands in 1980 and 1989.

The two charts in Exhibit 20.1 clearly show the loss of brand X's market share, apparently to brand B, between 1980 and 1989. Table 20.1 also offers the same insight, but it does not do so as vividly as Exhibit 20.1. While a single pie chart, by itself, can be quite revealing, a comparison of pie charts over time (e.g., as in Exhibit 20.1) or a comparison of related pie charts at the same point in time (e.g., pie charts showing the relative contributions of a firm's product lines to its sales and profits) can be even more insightful.

Using different shadings or colors for the various slices of a pie chart can improve its effectiveness. Also, limiting the number of slices to about six or seven is preferable. Too many slices will clutter the chart and diminish its visual impact. When a total quantity is made up of numerous components, one should lump minor components (i.e., those making relatively small individual contributions to the total) into an "Other" category to avoid overcrowding the chart.

Constructing a pie chart may not be helpful when the total is made up of numerous components, each contributing about equally to the total. For instance, when a firm's marketing budget is allocated to 20

Exhibit 20.1 *Pie Charts of Market Shares in 1980 and 1989*



different categories, each accounting for about 4% to 6% of the budget, a pie chart is likely to be no better than a table showing the allocations.

In short, pie charts will be most effective when the number of components is relatively small and the relative sizes of those components are dissimilar. Constructing pie charts under other circumstances may merely add to the length of a report without improving its effectiveness.

Line Charts

Line Chart *A two-dimensional graph typically used to show movements in one or more items over time.*

A ***line chart*** is a two-dimensional graph and is typically used to show movements in one or more items over time. The horizontal axis is customarily the time axis, and the vertical axis represents values of the items. Exhibit 20.2 is a line chart showing changes in market shares of the toothpaste brands during the period 1980 to 1989.

Exhibit 20.2 summarizes the data in Table 20.1 in a very revealing fashion. A glance at Exhibit 20.2 offers key insights (e.g., brand X's share has been steadily slipping, in almost symmetric contrast to brand B's increasing share) that only a very careful scrutiny of Table 20.1 will reveal.

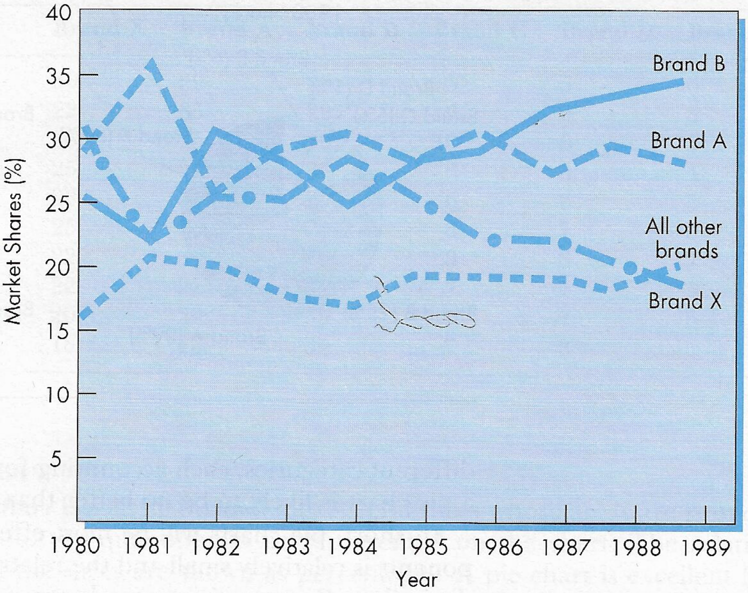
For a line chart to be most effective, the trends corresponding to different items must be shown by lines of different color or of different form, as in Exhibit 20.2. And just as in the case of pie charts, one should not include more than a few items on a line chart. Too many lines on one chart will give it a crowded appearance and cause confusion.

Stratum Chart *A two-dimensional graph with time along the horizontal axis and containing a series of distinct strata, or layers, stacked in the vertical direction. Each stratum represents a particular item, and at any given time the width of the stratum indicates the relative magnitude of the item at that time.*

Stratum Charts

A ***stratum chart*** is also a two-dimensional graph with time along the horizontal axis and values of the items to be plotted along the vertical axis. The area of the graph is divided into several horizontal layers, or strata, one corresponding to each item. At any given time the width of

Exhibit 20.2 *Line Chart of Market Shares from 1980 to 1989*



each stratum represents the relative magnitude of the corresponding item at that time. Exhibit 20.3 is a stratum chart summarizing the data in Table 20.1.

The lowest stratum in Exhibit 20.3 corresponds to brand X. The line defining the upper boundary of this stratum represents the same market share values as the brand X line in Exhibit 20.2. The second lowest stratum in Exhibit 20.3 corresponds to brand A. The line defining the upper boundary of this stratum represents the *sum* of the market shares of brands X and A. Similarly, the upper boundary of the third lowest stratum represents the combined share of brands X, A, and B.

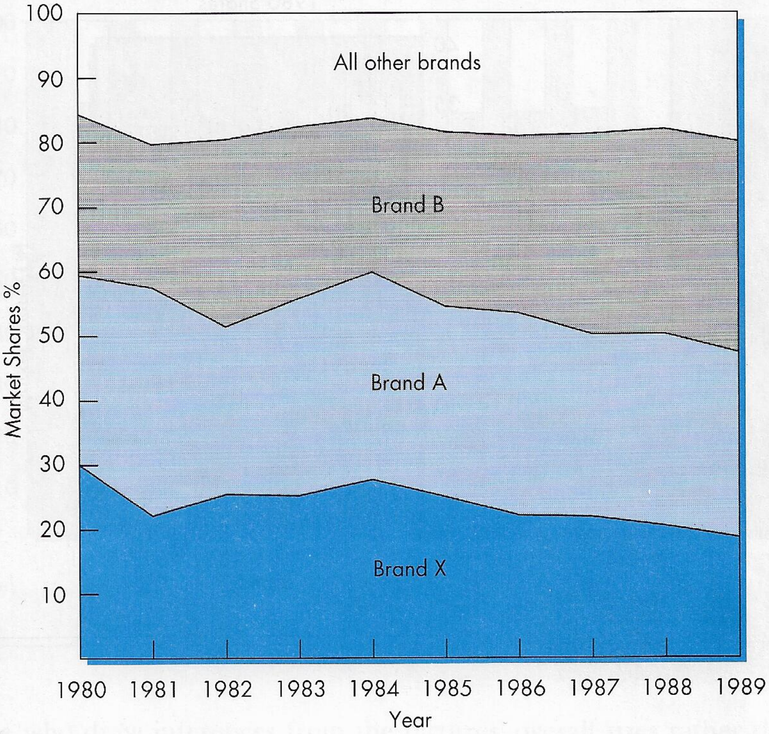
Thus a stratum chart is a *cumulative line chart*. Also, its information content is similar to that of a series of pie charts constructed over a period of time. That is, it tracks the changes in relative magnitudes of the items over time. Just as with pie and line charts, having a fairly small number of items (strata) and making the strata distinct by using different colors or shadings are critical for a stratum chart's effectiveness.

Bar Chart *A chart consisting of a series of bars (of equal thickness) whose heights (or lengths, if the bars are drawn horizontally) represent values of items.*

Bar Charts

A ***bar chart*** as its name implies, consists of a series of bars (of equal thickness) whose heights (or lengths, if the bars are drawn horizontally)

Exhibit 20.3 *Stratum Chart of Market Shares from 1980 to 1989*

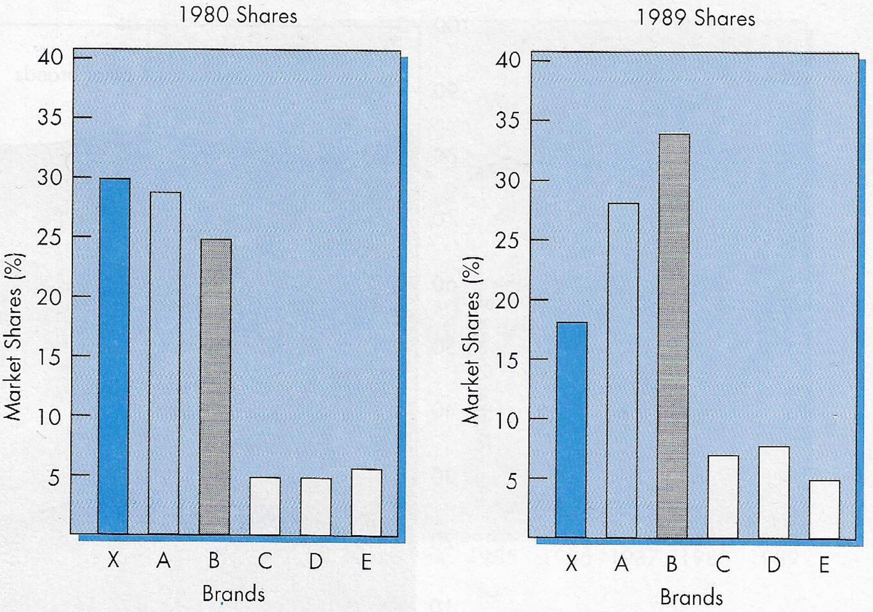


represent values of the items. Exhibit 20.4 contains two bar charts showing the market shares of toothpaste brands in 1980 and 1989.

The information conveyed by Exhibit 20.4 is the same as that conveyed by Exhibit 20.1. How, then, does one decide whether to use a pie chart or a bar chart? There is no cut-and-dried answer to this question. A pie chart is preferable, however, when the values to be pictured are *relative* magnitudes (e.g., market shares, budget shares, and other types of percentages), since the concept of slicing up a pie seems intuitively appropriate for representing relative shares. Bar charts are appropriate for depicting *actual* or *absolute* magnitudes (e.g., sales of different product lines, expenditures on different marketing activities). The heights (or lengths) of the bars can be made to correspond to the absolute magnitudes of items; yet the areas occupied by the bars will provide a visual picture of relative magnitudes of the items.

Bar charts can also be used for showing changes over time by, for example, plotting time along the horizontal axis, constructing one ver-

Exhibit 20.4 *Bar Charts of Market Shares in 1980 and 1989*



tical bar at each point in time, and dividing each bar into segments so that their heights correspond to the values of the items at that point in time. The visual impact of such a time series bar chart is similar to that of a stratum chart. Exhibit 20.5 shows a bar chart that conveys the same information as the stratum chart in Exhibit 20.3.

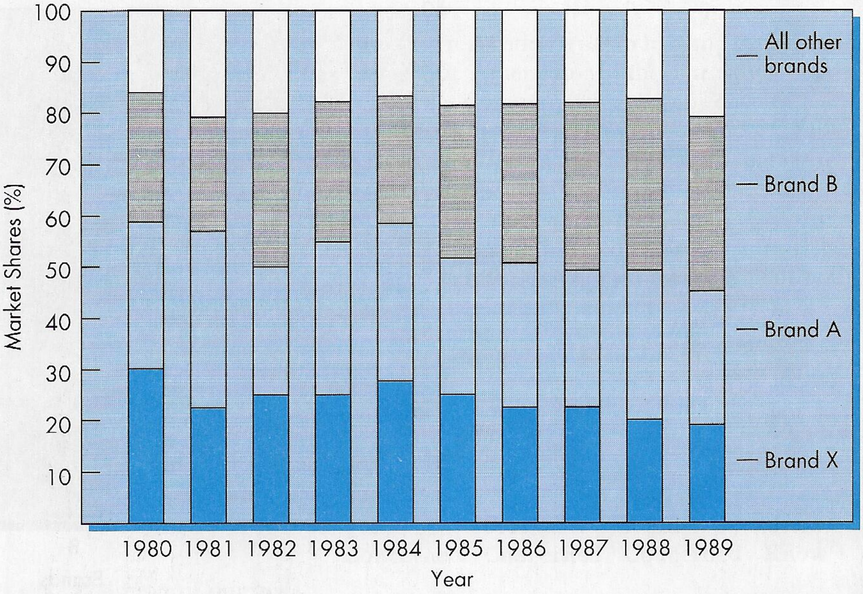
Pictograms

Pictogram A special type of bar chart that uses pictures of the items instead of the customary rectangular bars.

A ***pictogram*** is a special type of bar chart. It uses pictures of the items instead of the customary rectangular bars. The heights of the pictures correspond to the values of the items. A pictogram of the 1989 shares of the six toothpaste brands is shown in Exhibit 20.6.

The apparent advantage of a pictogram is that it is more appealing and interesting than a traditional bar chart. However, pictograms run the risk of misleading the audience through visual exaggeration. For instance, consider the two tubes of toothpaste representing brands A and C in Exhibit 20.6. Consistent with the fact that brand A's share is four times as much as brand C's share in 1989 (28% versus 7%), the brand A tube is four times as tall as the brand C tube. But the *area* occupied by the brand A tube is over six times as large as the area occupied by the brand C tube. The point is that a pictogram may mislead

Exhibit 20.5 *Bar Chart of Market Shares from 1980 to 1989*



those who draw inferences from the pictures' overall sizes rather than just their heights.[12]

Therefore a report using pictograms should ensure that the pictures do not mislead. Exhibit 20.6, for example, can be modified by representing the bars as stacks of *equal-sized* toothpaste tubes and varying the number of tubes in each stack to reflect different quantities. In this way the widths of the pictorial bars will be the same and only their heights will vary. Such a modified pictogram is shown in Exhibit 20.7.

Graphical Representation of One-Way and Two-Way Tabulations

Potential applications of the various charts described above go beyond the representation of market shares and trend lines. Indeed, several of those charts can be used to pictorially summarize information contained in one-way and two-way tabulations which are perhaps the most widely used procedures to analyze survey data in practical research projects. Research in Use 20.2 describes how one firm used a series of bar charts to visually, and effectively, represent the response distributions (i.e., one-

---

[12] An interesting discussion of the pitfalls of pictograms, with several good examples, is given in Darrell Huff, *How to Lie with Statistics* (New York: Norton, 1954), Chap. 6, pp. 66–73.

Exhibit 20.6 *Pictogram of Market Shares in 1989*

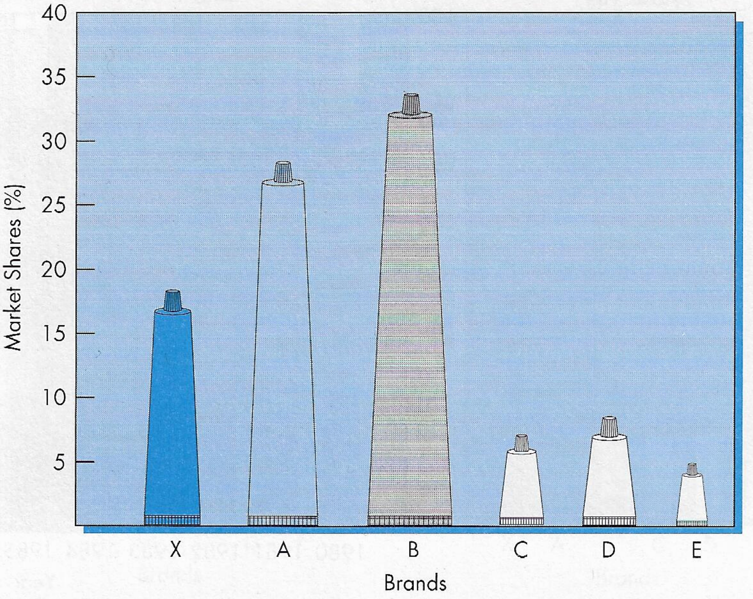
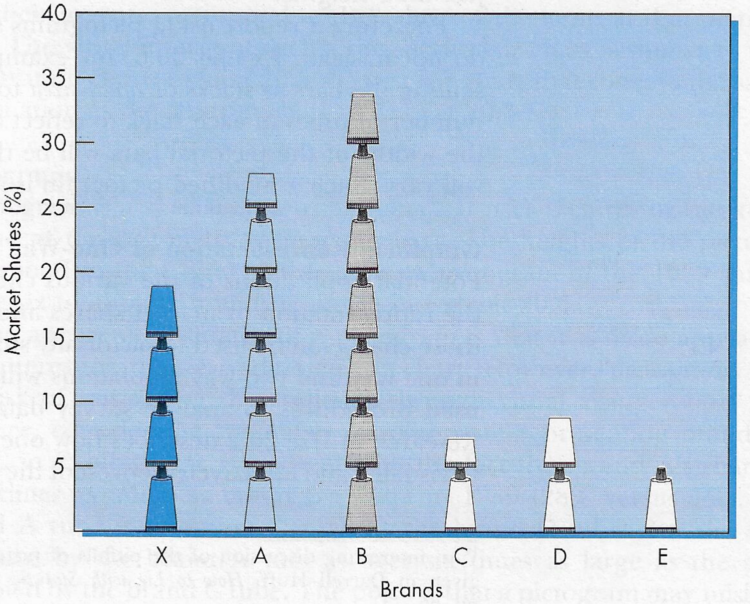


Exhibit 20.7 *Modified Pictogram of Market Shares in 1989*



Research in Use 20.2 *The Roper Organization*

Graphical Representation of One-Way Tabulations

The Roper Organization, a leading marketing research company, conducted a personal-interview survey of American youth. The study was commissioned by the Warner-Lambert Company with the objective of ascertaining young people's perception of their families, their schools, and the major social issues that confront them. A total of 1000 youth, between the ages of 8 and 17, participated in the study. The report of the study's findings included as an appendix a questionnaire with a summary of the responses (similar to the illustrative excerpts presented in Research in Use 20.1). The following is a sample section from that questionnaire (the numbers are percentages of respondents checking the answer categories).

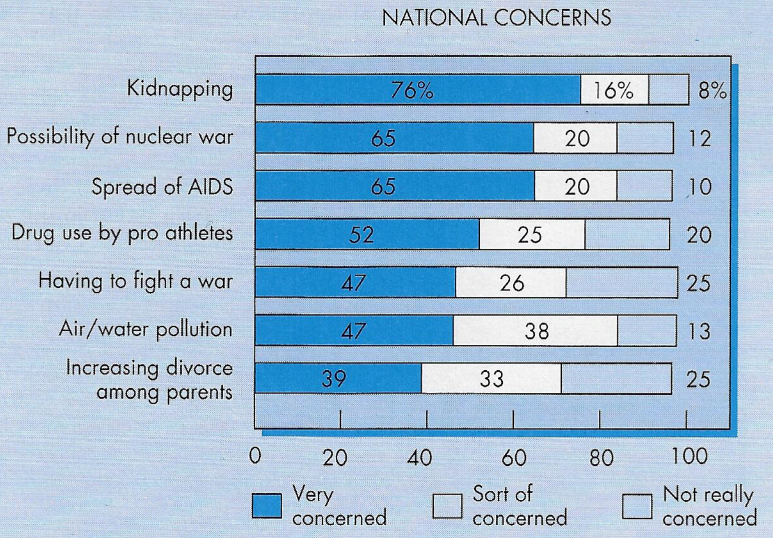
> Here is a list of things some young people have told us they are concerned about today. (HAND RESPONDENT CARD) Would you please read down that list and for each thing tell me if it is something you personally are very concerned about, sort of concerned about, or not really concerned about at all? First, (*read item*). (ASK ABOUT EACH)

	Very Concerned	**Sort of Concerned**	**Not Really Concerned**	**Don't Know**
a. Pollution of our air and water	47	38	13	2
b. The use of drugs by professional athletes	52	25	20	3
c. The possibility of nuclear war	65	20	12	3
d. The spread of the disease called AIDS	65	20	10	5
e. The increasing number of divorces among parents	39	33	25	3
f. The possibility that you may someday have to fight in a war	47	26	25	2
g. The kidnapping of children and teenagers	76	16	8	1

Each row of numbers in this section is essentially a one-way tabulation of responses pertaining to the issue in question. The discussion of these responses in the body of the report was accompanied by the following bar chart.

(Continued)

Research in Use 20.2 *Continued*



This bar chart, wherein each bar is a pictorial summary of a one-way table, quickly and effectively communicates the findings to the reader. Notice that the issues are ordered differently in the bar chart than in the questionnaire. The ordering of the issues on the bar chart (from those of most to least concern) enhances the chart's effectiveness.

Source: "The American Chicle Youth Poll," a report published by the Warner-Lambert Company, 201 Tabor Road, Morris Plains, NJ 07950. Used by permission.

way tabulation of responses) for a number of questions with categorical responses.

Bar charts can also be used to clearly and quickly communicate information contained in two-way tabulations. To illustrate, consider the two-way tabulation shown in Table 20.2. This table shows a cross-tabulation of responses (obtained from a survey of 1000 consumers) pertaining to two variables: consumers' age (coded as <30, 30–50, or >50) and extent of consumption of a certain product (coded as heavy, medium, or light). While the nature of the association between these two variables can be inferred from the pattern of percentages in the table, it can be conveyed much more efficiently and effectively by displaying the percentages on a chart such as the one in Exhibit 20.8. There is clearly a strong association between age and extent of consumption. The under-30 group has the heaviest consumption and the 30–50 group has the lightest consumption, with the consumption level of the over-50 group falling in between.

Table 20.2 *Two-Way Tabulation of Age Versus Extent of Consumption*

	Extent of Consumption			
Age	**Heavy**	**Medium**	**Light**	**Total**
< 30	125 (50%)	65 (26%)	60 (24%)	250 (100%)
30–50	60 (15%)	80 (20%)	260 (65%)	400 (100%)
> 50	65 (18%)	205 (59%)	80 (23%)	350 (100%)
Total	250 (25%)	350 (35%)	400 (40%)	1000 (100%)

Note: All percentages are based on row totals; some percentages have been adjusted slightly to make them add to 100%

A bar chart similar to Exhibit 20.8 is not the only approach to graphically representing two-way tabulations. As was mentioned at the beginning of this section, one's imagination is the only limit on how to transform tabular information into pictures. For instance, instead of the four clusters of three bars each shown in Exhibit 20.8, one can construct four pie charts with each containing three slices.

Computer Graphics

The proliferation of personal computers and the availability of a variety of computer graphics software packages have greatly enhanced the options available for presenting research findings as graphical illustrations.[13] Many computer graphics programs can analyze data stored in memory and present the results in a variety of pictorial formats. The user can peruse visual displays of the resulting pictures and obtain hard copies of the most effective ones. This capability should simplify the task of determining the best way to graphically portray a study's findings. In fact, computer experts predict that in the near future "some graphics software will feature artificial intelligence that will be able to select automatically the most effective graph for a particular set of data."[14]

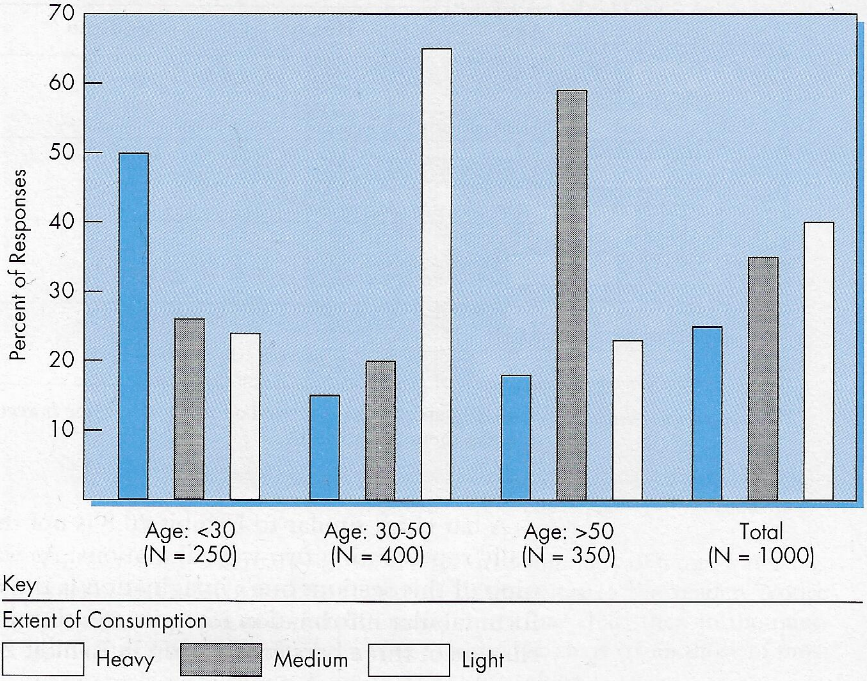
Machines are also available now to create color photographs, prints, slides, and overhead transparencies of computer-generated charts or graphics. These machines are an integral part of, or can be attached to,

---

[13] For a listing of available computer graphics software packages, see, for example, "1989 Directory of Software for Marketing," *Marketing News*, April 24, 1989, pp. 12–24.

[14] "Management Warms Up to Computer Graphics," *BusinessWeek*, August 13, 1984, p. 101.

Exhibit 20.8 *Bar Chart of Information in Table 20.2*



many personal computers. Therefore researchers can easily coordinate their written reports with their oral presentations by creating colorful overhead transparencies and/or slides of key charts included in the report. The upshot is that written reports and oral presentations without graphical illustrations are more an exception than the rule today and are bound to become even more so in the future.

The ease of generating graphical illustrations of a study's results does not mean that a diagram or chart should be produced for every result. While graphical illustrations can make a report interesting and add to its clarity, *too many* of them may clutter up a report and annoy the audience. The number of illustrations to use and the appropriate formats for them should therefore be carefully determined to avoid confusing, meaningless, or misleading illustrations.[15]

---

[15] For a discussion of situations where certain graphical illustrations are more appropriate than others, as well as the pros and cons of each, see Mary E. Spear, *Practical Charting Techniques* (New York: McGraw-Hill, 1969). See also Anderson, *Business Communication*, Chap. 15.

ORAL PRESENTATIONS

Oral presentations, like written reports, must also be SIMPLE: short, interesting, methodical, precise, lucid, and error-free. However, making an effective oral presentation is in some ways more difficult than writing a good report because of the direct interaction with the audience. Any sign of faltering during an oral presentation will make an unfavorable impression on the audience and may lower the self-confidence of the presenter. An effective oral presentation requires meticulous preparation of what will be said and how it will be said. It also requires planning for contingencies like a breakdown of visual equipment or a series of tough questions from the audience. Even a thirty-minute presentation may therefore need many hours (or days) of preparation. Carefully planning and rehearsing an oral presentation is critical to its effectiveness.

Details for putting together and delivering a good oral report can be found in several excellent books.[16] However, the following three general tasks, covered while discussing written reports, are especially important for the success of oral presentations:

1. Researching the audience—knowing who will be in the audience, understanding their backgrounds and information needs, and anticipating the questions they may ask.
2. Choosing the main points of the study to be covered during the presentation, while being careful not to select too many main points.
3. Making good use of visual aids—flip charts, transparencies, slides, and so on—to improve presentation clarity and maintain audience interest.

Concerning the third point, given the proliferation of technology and equipment for creating visual aids, such aids are expected by audiences and used by presenters in virtually all oral presentations. Indiscriminate use of visual aids is not likely to be of much help, however. They must be chosen and sequenced with care, with each one containing a simple, relevant message. Exhibit 20.9 provides a number of suggestions for the proper use of slides. Many of the suggestions are also relevant for ensuring the effectiveness of other forms of visual aids.

---

[16] A succinct discussion is given in Timm, *Managerial Communication,* Chap. 7; see also Leonard F. Meuse, Jr., *Succeeding at Business and Technical Presentations* (New York: Wiley, 1988), Chap. 10.

Exhibit 20.9 *Guidelines for Effective Use of Slides*

WORD SLIDES

- Keep them brief; use key words only.
- Use bullets and color to highlight key points.
- Break up the information to make a series of slides (a progressive or "build" series). Use color to show the new line added to each slide.

TABULAR SLIDES

- Use to show lists.
- Keep items brief as possible; arrange them to fill the slide area so the type can be as large as possible.

BOX CHARTS

- Use for organization charts, flow charts.
- Simplify to keep them legible.
- Break up complex charts into a series. (Show flow chart divided by time periods; show organization chart with the overall chart and departmental "close up.")

BAR CHARTS

- Use them for data arranged in segments (by month, year, etc.)
- Choose vertical or horizontal bars (both within horizontal slide format).
- Add drop shadow for dimensional bars.
- Show complex facts clearly by using multiple or segmented bars.
- Divide extensive data into a progressive-disclosure series.

PIE CHARTS

- Use them to emphasize the relationship of the parts to the whole.
- Select single pie or double pie.
- Consider options such as drop shadow for dimensional effect, pulled-out slices, etc.
- Arrange the slices to make your point most effectively.
- Divide the slice into a series if that improves effectiveness.

LINE GRAPHS AND AREA GRAPHS

- Use these to display trends or continuous data.
- Decide whether line graph or area graph shows your point better.
- Select baseline and scale for maximum effectiveness.
- Use callouts to identify key points in graph.
- Divide extensive data into a series of graphs.

(Continued)

Exhibit 20.9 *Continued*

OTHER GRAPHICS

- Computer technology can be used to produce many additional kinds of graphics and illustrations.
- Since these can be stored in computer memory and recalled for reuse, they can be very economical.
- Logos and illustrations can be used in subdued colors in the background, as "watermarks"—an effective way to add visual interest and continuity to a presentation.
- Any previously created slides can be recalled from memory and combined to make new slides.

Source: From Leslie Blumberg, "For Graphic Presentations, Managers Focus on Slides," *Data Management,* May 1983, p. 23. Copyright and reprint permission granted by Data Processing Management Association. All rights reserved.

SUMMARY

Communicating the results of a research project to decision makers is the last step of the research process. It is by no means the least important step, however. Indeed, it may be the *most* important step in that the practical worth of a project is closely linked to how effectively its findings are communicated. The foremost step in communicating the results of a research project is to gain a good understanding of the audience and its information needs. Lack of awareness of the audience and their concerns is frequently why many sound research projects do not have the impact they should.

The most common vehicle for communicating research findings is the written report. To be effective, a written report must be SIMPLE; *s*hort, *i*nteresting, *m*ethodical, *p*recise, *l*ucid, and *e*rror-free. Although these six features are interrelated, they each have certain distinct requirements:

- *Short.* A report should discuss only those aspects of a study that are relevant and important from the reader's standpoint. Other aspects such as methodological details, which may fascinate researchers but frustrate readers, must be omitted or only briefly stated.
- *Interesting.* The style, format, and contents of a report must attract and maintain the attention of readers.
- *Methodical.* The sections of a report and their contents must be chosen with the readers in mind and arranged in a logical sequence. While there is no standard research report format, the outline suggested in the chapter should be helpful.
- *Precise.* A report must be clear and comprehensive. Clarity is a function of writing style and format. Comprehensiveness implies that the report must contain enough detail to provide an accurate, complete

picture of the study; the report should acknowledge key limitations and how they may affect interpretation of the results.

- *Lucid.* The report must be written in language that is not likely to be foreign to its readers. It should use simple words and sentences, as well as appropriate tables and charts, so as to communicate its contents clearly and quickly.
- *Error-free.* Even seemingly minor arithmetical, grammatical, or typographical errors in a report may annoy its readers a great deal. Sufficient time and effort must be devoted to catching and correcting such errors before finalizing the report.

A variety of graphical illustrations—such as pie charts, line charts, stratum charts, bar charts, and pictograms—can be employed to enhance the clarity and effectiveness of a report. However, the number and types of graphical illustrations to use must be determined with care. Haphazard use of illustrations may merely add to a report's bulk; worse still, it may confuse or mislead readers.

In addition to submitting written reports, researchers often make oral presentations of their findings. Many of the guidelines for preparing good written reports apply to making effective oral presentations as well. Understanding the audience, distilling from the study the main messages to be conveyed, and making good use of appropriate visual aids are particularly germane for oral presentations. For a successful oral presentation, adequate time and effort must be spent on planning and rehearsing it.

KEY TERMS

SIMPLE:
*S*hort
*I*nteresting
*M*ethodical
*P*recise
*L*ucid
*E*rror-free

Pie chart
Line chart
Stratum chart
Bar chart
Pictogram

QUESTIONS

1. The last step in the research process—communicating research results—may well be the most important step. What are the major reasons?
2. Two desirable features of a written report are that it be short and that it be precise. Is there likely to be any conflict between the two? If so, how would you reconcile such a conflict? If not, why not?
3. "To have an impact on readers, a research report must be humorous and entertaining." Discuss this statement.

4. Why is the executive summary such a critical component of a research report?

5. Recast the following statements in less technical language:
   a. "On the basis of our survey, the 95% confidence interval for the proportion of households with VCRs is .4 ± .05."
   b. "The regression of sales (dependent variable) on advertising expenditures (independent variable) produced an R^2-value of .9, which was statistically significant ($F = 55.8$, $p < .01$)."

6. Go through a business publication like *BusinessWeek* or *Fortune* and pick out any three graphical illustrations. State what type of chart or graph each is, and critically evaluate it—that is, identify its positive and negative features.

7. Using the data in Table 20.1, compute the rate of change in market shares for brands X, A, and B from year to year (i.e., 1980 to 1981, 1981 to 1982, etc.). Diagram these changes by using a line chart and a bar chart. Which of these two charts is more effective in your opinion? Why?

8. Can a pie chart or a stratum chart be used to depict the market share changes computed in Question 7? Why or why not?

9. Ms. Smith, a researcher in XYZ Company, is delighted because management just informed her that she need only make a brief oral presentation of a study she recently completed; management does not want her to write a written report. "Thank goodness I don't have to go through the torture of writing up my findings!" she thinks. What advice would you give to Ms. Smith?

Appendices

A *NPD Data Set and Analysis Exercises*

B *Areas Under the Standard Normal Curve*

C *Chi-square Distribution*

D *Student's **t**-Distribution*

E *Critical Values of the Sample Correlation Coefficient **r***

F *Values of $F_{.05}$*

APPENDIX A

NPD Data Set and Analysis Exercises

INTRODUCTION

This appendix provides and discusses a data set for illustrating a number of analysis concepts and techniques covered in this book, particularly in Chapters 16, 17, 18, and 19. The data set contains real-life data culled from purchase diaries filled out over a period of one year by a representative sample of 100 families. These families are from a consumer panel maintained by the NPD Research, Inc., of New York. A description of NPD's consumer panel and a sample page of NPD's purchase diary are contained in the appendix to Chapter 5.

The data set consists of 200 data records, two per family. The data on each family include certain demographic data and purchase data pertaining to four product categories: toilet tissue, frosting, cereal, and mayonnaise. The next section presents a more detailed description of the data set.

VARIABLES INCLUDED IN THE DATA SET

Demographic Data

Data are available for the 100 families on the following 10 demographic characteristics (labeled as variables D1 through D10 for future reference):

1. Education of female head (D1)
2. Employment status of female head (D2)
3. Pet ownership (D3)
4. Type of residence (D4)
5. Residence ownership (D5)
6. Age and presence of children (D6)
7. Race (D7)

The author thanks NPD Research Inc., of New York for providing the raw data and instructions necessary for constructing the data set included here.

8. Family income (D8)
9. Age of male head (D9)
10. Education of male head (D10)

The demographic data are in columns 4 through 14 of each family's first record. Table 1 indicates the exact location of data on each variable. Table 2 explains the numerical scheme used in coding the data on each variable. Table 3 contains the data set. All three tables are at the end of this appendix.

Purchase Data

The purchase data for each of the four products span a period of 12 months: from July of one year to June of the next year. However, so that the data set's size is not too large, data pertaining to only the *first* and the *last* purchase during the 12-month period are reported. For each purchase within each product category, the purchase-related variables included in the data set are as follows:

1. Brand code
2. Quantity purchased (number of units)
3. Total price paid (for the entire quantity purchased)
4. Regular purchase or special purchase (i.e., purchase made when a special promotion was in effect)
5. Month of purchase

For the cereal product category, data are available on two additional variables: manufacturer's name and type of cereal.

The labels shown in Tables 1 and 2 for the purchase-related variables each consist of three characters. The first character represents the product category (T for toilet tissue, F for frosting, C for cereal, M for mayonnaise), the second character represents the purchase occasion (A for first purchase and B for last purchase), and the third character represents the variable itself (1 for brand code, 2 for quantity purchased, 3 for total price paid, 4 for regular purchase or special purchase, 5 for month of purchase, M for manufacturer's name, T for type of cereal; the last two symbols are relevant only to the cereal category).

The data set also includes the total number of purchases made by each family, within each product category, during the 12-month period. The variables denoting the total number of purchases are labeled in Tables 1 and 2 as TNP for toilet tissue, FNP for frosting, CNP for cereal, and MNP for mayonnaise.

AN ILLUSTRATION

To better understand the manner in which the data set (Table 3) is structured, let us infer from it the demographic characteristics and the purchase behavior of family 003. Specifically, let us interpret the fifth and sixth rows in the data set with the help of the coding guidelines presented in Tables 1 and 2.

Family 003 has the following demographic characteristics: The family's female head has had some college education; she is not employed. The family has no dog or cat, lives in a one-family house, owns the house, has children in all three age groups (under 6, 6–12, and 13–17), is white, and has an annual income of \$10,000–\$11,999. The family's male head is 55–64 years old; he has had some college education.

The family's first purchase of toilet tissue was brand 30. The family bought one unit of the brand, paid $0.49 for the unit, and made the purchase at a special price. This purchase occurred in month 02 (August of the first year). The family's last purchase of toilet tissue was brand 05. The family bought two units of this brand, paid a total of $1.76 for the two units, and made the purchase at a special price. This purchase occurred in month 12 (June of the second year). This family made a total of 10 toilet tissue purchases during the year.

Family 003 did not make any purchases of frosting during the year. (The rest of this family's first data record, beginning with column 43, is blank except for the zeros corresponding to FNP.)

The family's first purchase of cereal was brand 90, a private-label brand and a regular type. The family bought one unit of the brand, paid $1.10 for it, and made the purchase at regular price. This purchase occurred in month 02. The family's last purchase of cereal was brand 07, a Nabisco brand and a regular type. The family bought one unit of the brand, paid $0.93 for it, and made the purchase at a special price. This purchase occurred in month 12. This family made a total of 17 cereal purchases during the year.

The family's first purchase of mayonnaise was brand 003. The family bought one unit of the brand, paid $1.12 for it, and made the purchase at a special price. This purchase occurred in month 01. The family's last purchase of mayonnaise was brand 032. The family bought one unit of the brand, paid $0.69 for it, and made the purchase at a special price. This purchase occurred in month 12. The family made a total of two mayonnaise purchases during the year.

The next section of this appendix presents a number of exercises involving the NPD data set. These exercises require applying analysis concepts and techniques covered in Part Four of the text. The exercises are numbered sequentially but grouped according to the chapters to which they relate. Availability of the data set on computer and access to a standard statistical package such as SPSS or SAS are necessary for completing most of the exercises.

EXERCISES

Exercises for Chapter 16

1. Indicate the measurement level (i.e., nominal, ordinal, interval, or ratio scale) for each variable listed in Table 2.
2. Answer the following questions after creating the necessary transformed variables and analyzing the data on those variables.
   a. What proportion of the families who purchased toilet tissue at least twice during the year bought the *same* brand at the first and the last purchases?
   b. What proportion of the families who purchased mayonnaise at least twice during the year bought the *same* brand at the first and last purchases?
   c. What inference(s) can you make about the degree of brand loyalty for toilet tissue and mayonnaise? Explain your answer.
3. For each of the four product categories, set up the following transformed variables:

$$\begin{aligned}\text{purchase duration} &= (\text{month of last purchase})\\ &\qquad - (\text{month of first purchase}) + (1),\\ \text{purchase intensity} &= (\text{total number of purchases during the year})\\ &\qquad \div (\text{purchase duration}).\end{aligned}$$

Compute the average purchase intensity (across all families) for each product category. What insights emerge from comparing these average purchase intensities?

4. In what proportion of the sample families having male and female heads do *both* heads have at least some college education?

5. Determine and interpret the appropriate measure of central tendency for each of the following variables: D3, D6, D8, TA2, TB2, FA4, FB4, CA4, CB4, TNP, FNP, CNP, and MNP.

6. Construct, compare, and interpret the one-way tables for variables in each of the following sets of variables:
a. D1 and D10.
b. TA4, TB4, MA4, and MB4.
c. TA5, FA5, CA5, and MA5.

7. Answer the following questions after constructing and examining the appropriate two-way tables.
a. Is there an association between whether the male family head has had at least some college education and whether the family income is at least $25,000?
b. Is there an association between whether a family has children under 18 years of age and whether the family owns a dog or a cat?
c. Is there an association between whether family income is at least $25,000 and whether a purchase is made at regular price for the first purchase of (i) toilet tissue, (ii) frosting, (iii) cereal, and (iv) mayonnaise? What inferences can you draw by comparing the results across the four product categories?

Exercises for Chapter 17

8. A firm manufacturing several brands of frosting is wondering whether to run a special advertising campaign to improve the frequency of purchase for the product. The firm wants to run the campaign only if it is reasonably sure (i.e., 95% confident) that the average family makes less than two frosting purchases during the year. On the basis of the NPD sample data on frosting purchases, should the firm run the campaign?

9. Considering only families that purchased a regular, presweetened, or nutritional cereal during the first purchase (i.e., the first three categories of variable CAT in the NPD data set), test the following null hypothesis at a significance level of .01: Family preferences for the three types of cereals are identical.

10. Assume that the income distribution of families in the United States is as follows.

Less than $10,000	15%
$10,000–$14,999	20%
$15,000–$24,999	25%
$25,000–$34,999	25%
$35,000 or more	15%
	100%

Can we be 95% confident that the NPD sample of families is representative of all United States families with respect to income?

11. Using the NPD data, test the null hypothesis that at least 30% of all families buy mayonnaise less than three times a year. (Assume an alpha value of .01.)

12. Is the mean number of cereal purchases per year for families with children under 18 significantly different from that for families with no children under 18? (Assume a significance level of .05.)

13. Is the mean number of units of toilet tissue bought during the first purchase significantly different from the mean number of units of toilet tissue bought during the last purchase, assuming a significance level of .05?

14. Are the proportions of families buying toilet tissue at regular price during the first purchase significantly different for families with an income of less than $25,000 and for those with an income of at least $25,000? (Assume a significance level of .1.)

Exercises for Chapter 18

15. Conduct a chi-square contingency test for each of the tables you set up under Exercise 7, assuming a significance level of .05.

16. Using a significance level of .1, repeat the chi-square contingency test corresponding to the first purchase of toilet tissue that you conducted under Exercise 15. Compare your inference from this test with that in Exercise 14. Are the two inferences alike? Explain your answer.

17. For which of the various chi-square contingency tests you conducted under Exercise 15 would computing the contingency coefficient and the Cramer's *V*-statistic be meaningful? Compute these values where appropriate and interpret them.

18. What is the nature and the degree of association between the education of the female head and that of the male head of families included in the NPD sample? Is this association statistically significant at the .05 significance level?

19. A firm's product line includes toilet tissue, frosting, cereal, and mayonnaise. This firm wants to ascertain the nature and the extent of association between the purchase frequency (i.e., total number of purchases per year) for each product and the purchase frequency for each of the other three products. To achieve the firm's information objective, perform appropriate analyses of the data on the TNP, FNP, CNP, and MNP variables in the NPD daa set, and interpret the results.

20. Repeat Exercise 19 by analyzing data on *purchase intensity* (as defined in Exercise 3) rather than on purchase frequency. Do the results and inferences differ from those obtained in Exercise 19? Explain your answer.

21. For the first purchase of each product category in the NPD data set, perform the following simple regression analysis:

$$\text{quantity purchased} = a + b\ (\text{price per unit}).$$

(Note that you first need to transform the "total price paid" variable in the NPD data set so as to obtain values for the independent variable in this equation.) What inferences stem from your analyses concerning the relative sensitivity of quantity purchased to price across the four product categories? How meaningful and trustworthy are those inferences? Why?

22. For each of the four product categories, conduct a multiple regression anaysis with *purchase duration* (as defined in Exercise 3) as the dependent variable and with total number of purchases and family income as the independent variables. (Recode the 14 income categories as 1 through 14 and assume the resulting data are metric-scaled for illustrative purposes. Use the recorded income variable in the regression analysis.) Answer the following questions for each multiple regression equation:

a. Is the regression equation trustworthy and useful? Explain.
b. Are the regression coefficients statistically significant?
c. Do the regression coefficients have meaningful interpretations? Is the problem of multicollinearity present? If so, is it serious? Explain your answers.

Exercises for Chapter 19

23. Create a transformed income variable with three categories as follows:

Low	$14,999 or less
Medium	$15,000 to $29,999
High	$30,000 or more

Considering the low, the medium, and the high categories as three treatments, conduct a one-way analysis of variance to ascertain whether differences exist across the three income groups in terms of the purchase frequency (i.e., total number of purchases during the year) for toilet tissue, frosting, cereal, and mayonnaise. Discuss the results.

24. Using the data pertaining to the first purchase of toilet tissue, perform an appropriate one-way analysis of variance to ascertain whether the quantity purchased (TA2) varies significantly depending on whether the purchase was made at regular price or at a special price (TA4).

25. Repeat Exercise 24 by treating TA4 as a dummy variable and conducting a simple regression analysis. Interpret the results and compare them with those in Exercise 24.

Table 1 *Location of Demographic and Purchase Data on Each Family's Data Records*

Column Number	First Record	Second Record
01–03	Family code number	Blank
	Demographic Data	
04	Education of female head (D1)	Blank
05	Employment status of female head (D2)	Blank
06	Pet ownership (D3)	Blank
07	Type of residence (D4)	Blank
08	Residence ownership (D5)	Blank
09	Age and presence of children (D6)	Blank
10	Race (D7)	Blank
11–12	Family income (D8)	Blank
13	Age of male head (D9)	Blank
14	Education of male head (D10)	Blank
	Toilet Tissue Purchase Data	**Cereal Purchase Data**
	First purchase of the year:	First purchase of the year:
15–16	Brand code (TA1)	Brand code (CA1)
17	Blank	Manufacturer's name (CAM)
18	Blank	Type of cereal (CAT)
19–20	Quantity purchased (TA2)	Quantity purchased (CA2)
21–24	Total price paid (TA3)	Total price paid (CA3)
25	Regular or special purchase (TA4)	Regular or special purchase (CA4)
26–27	Month of purchase (TA5)	Month of purchase (CA5)
	Last purchase of the year:	Last purchase of the year:
28–29	Brand code (TB1)	Brand code (CB1)
30	Blank	Manufacturer's name (CBM)
31	Blank	Type of cereal (CBT)
32–33	Quantity purchased (TB2)	Quantity purchased (CB2)
34–37	Total price paid (TB3)	Total price paid (CB3)
38	Regular or special purchase (TB4)	Regular or special purchase (CB4)
39–40	Month of purchase (TB5)	Month of purchase (CB5)
41–43	Total number of purchases during the year (TNP)	Total number of purchases during the year (CNP)
	Frosting Purchase Data	**Mayonnaise Purchase Data**
	First purchase of the year:	First purchase of the year:
44–46	Brand code (FA1)	Brand code (MA1)
47–48	Quantity purchased (FA2)	Quantity purchased (MA2)
49–52	Total price paid (FA3)	Total price paid (MA3)
53	Regular or special purchase (FA4)	Regular or special purchase (MA4)
54–55	Month of purchase (FA5)	Month of purchase (MA5)
	Last purchase of the year:	Last purchase of the year:
56–58	Brand code (FB1)	Brand code (MB1)
59–60	Quantity purchased (FB2)	Quantity purchased (MB2)
61–64	Total price paid (FB3)	Total price paid (MB3)
65	Regular or special purchase (FB4)	Regular or special purchase (MB4)
66–67	Month of purchase (FB5)	Month of purchase (MB5)
68–70	Total number of purchases during the year (FNP)	Total number of purchases during the year (MNP)

Table 2 *Numerical Coding Scheme for Data Set*

Variable Description	Variable Label(s)	Description of Numerical Codes Representing Variable
Education of female head	D1	1 Grade school
		2 Some high school
		3 Graduated high school
		4 Some college
		5 Graduated college
		6 Post graduate
		0 No female head
Employment status female head	D2	1 Employed under 30 hours
		2 Employed 30–34 hours
		3 Employed 35 or more hours
		8 Employed hours unknown
		9 Not employed
Pet ownership	D3	0 No dog or cat
		1 Dog only
		2 Cat only
		3 Both cat and dog
Type of residence	D4	1 One-family house
		2 Building for two families
		3 Building for three or more families
		4 Mobile home or trailer
		5 Other
Residence ownership	D5	1 Own
		2 Rent or lease
		3 Rent free
Age and presence of children	D6	1 Under 6 only
		2 6–12 only
		3 13–17 only
		4 Under 6 & 6–12
		5 Under 6 & 13–17
		6 6–12 & 13–17
		7 All three age-groups
		9 No children under 18
Race	D7	1 White
		2 Black
		3 Oriental
		4 Other

(Continued)

Table 2 *Continued*

Variable Description	Variable Label(s)	Description of Numerical Codes Representing Variable
Family income	D8	04 Under 8000 06 8000–9999 08 10,000–11,999 10 12,000–14,999 11 15,000–19,999 13 20,000–24,999 15 25,000–29,999 16 30,000–34,999 17 35,000–39,999 18 40,000–44,999 19 45,000–49,999 21 50,000–59,999 23 60,000–69,999 24 70,000 and over
Age of male head	D9	1 <25 2 25–29 3 30–34 4 35–39 5 40–44 6 45–49 7 50–54 8 55–64 0 65+ 0 No male head
Education of male head	D10	1 Grade school 2 Some high school 3 Graduated high school 4 Some college 5 Graduated college 6 Post graduate 0 No male head
Brand code	TA1, FA1, CA1, MA1, TB1, FB1, CB1, MB1	A two-digit code number for toilet tissue and cereal, and a three-digit code number for frosting and mayonnaise (brand names are confidential)
Quantity purchased	TA2, FA2, CA2, MA2, TB2, FB2, CB2, MB2	A two-digit number representing the actual number of units of the brand purchased
Total price paid	TA3, FA3, CA3, MA3, TB3, FB3, CB3, MB3	A four-digit number representing the actual price paid (in cents) for the total number of units of the brand purchased (e.g., a 0784 represents a total price of $7.84)

(Continued)

Table 2 *Concluded*

Variable Description	Variable Label(s)	Description of Numerical Codes Representing Variable
Regular or special purchase	TA4, FA4, CA4, MA4, TB4, FB4, CB4, MB4	1 Regular price 2 Special price 3 Free sample or gift
Month of purchase	TA5, FA5, CA5, MA5, TB5, FB5, CB5, MB5	A two-digit number from 01 to 12; 01 corresponds to July of one year and 12 corresponds to June of the next year
Manufacturer's name (relevant only for cereals)	CAM, CBM	1 Quaker 2 General Mills 3 Kellogg 4 Nabisco 5 Post 6 Ralston 7 Sunshine 8 Private label 9 All other
Type of cereal	CAT, CBT	1 Regular 2 Presweetened 3 Nutritional 4 Variety pack 5 Granola health food 6 Wheat germ
Total number of purchases during the year	TNP, FNP, CNP, MNP	A three-digit number representing the actual number of times the product category was purchased during the entire year

Note: For all variables, blank spaces in the data set represent missing values.

Table 3 *NPD Data Set*

Record Columns													
01–05	**06–10**	**11–15**	**16–20**	**21–25**	**26–30**	**31–35**	**36–40**	**41–45**	**46–50**	**51–55**	**56–60**	**61–65**	**66–70**
00129	02191	04940	3 02	01781	0903	0201	78111	00200	10101	39110			001
		2	73101	01231	01153	10101	29111	00800	30100	78101	00301	01491	12005
00249	01191	16850	5 01	00692	0310	0504	00207	002					000
								00000	10101	59112			001
00349	01171	08843	0 01	00492	0205	0201	76212	010					000
		9	08101	01101	02074	10100	93212	01700	30101	12201	03201	00692	12002
00441	01191	24850	3 01	01151	0103	0202	14212	01001	20101	39110			001
		1	33101	01951	02133	10101	87111	00402	40101	69107	02901	01152	12002
00529	03291	10920	5 01	01291	0139	0100	90212	027					000
		1	53101	00942	01			001					000
00643	01192	13820	5 01	01792	0105	0101	59212	01400	30101	75106			001
		0	24101	01591	02195	10101	69212	00600	30100	86202	01301	01291	05003
00731	31191	11830	1 02	01001	0101	0401	96112	025					000
		3	03101	01181	08743	10101	91108	00300	60101	11112			001
00831	11161	13540	3 01	00992	0101	0101	26212	02600	30100	84201	00101	00842	12021
		8	28201	01591	01023	20101	73212	10600	30100	99202			001
00939	01191	11921	8 02	01242	0218	0101	39212	01409	30201	00201	00101	00792	12012
		1	93201	01351	09015	20101	84212	00500	30100	84201	00301	01272	09004
01041	31191	18844	3 02	01502	0104	0201	58211	011					000
		0	74101	01221	07214	10101	58112	01100	10101	39101	00101	01492	12012
01153	21191	21661	8 06	06542	0318	0607	14111	00558	50100	69101	58501	00621	09002
		1	52101	01242	01152	10101	01212	02100	10101	89101	00301	01842	12005
01239	01161	10430	5 02	01582	0154	0101	89112	01354	50201	98202	00101	01172	09007
		1	35201	01092	01282	10302	67112	05604	50100	85202	52902	01381	12003
01319	02291	04008	5 01	00791	0339	0101	03112	005					000
		9	29101	00431	01175	10100	99112	02506	50101	09101	55701	01291	12011
01429	31291	04430	1 01	01691	0239	0201	78112	01401	20101	19109			001
		0	92301	01791	01868	20101	09112	04600	30100	89102	00301	01091	12007
01539	01191	11830	1 02	00901	0101	0200	39212	02000	30101	19204			001
		1	62201	01991	11415	20101	79112	00300	50100	29201	00101	2	12009
01649	01391	10920	1 02	00961	0101	0101	97112	010					000
								00003	20100	89111			001
01763	14 91	13968	5 01	00691	0203	0100	77112	031					000
		3	03101	01201	01125	10101	70112	02100	30100	97202	00501	00772	12008
01852	01161	10450	1 01	00381	0101	0100	39110	01000	10101	39104	00301	01491	11005
		0	81301	01991	01081	30101	99111	03500	30101	49104	00301	01091	11004
01939	11131	11810	5 02	01422	0103	0203	06212	02600	10100	79210			001
		0	62201	01471	01074	10101	48112	01400	30100	89202	00301	01262	10008

(Continued)

Table 3 *Continued*

Record Columns													
01–05	**06–10**	**11–15**	**16–20**	**21–25**	**26–30**	**31–35**	**36–40**	**41–45**	**46–50**	**51–55**	**56–60**	**61–65**	**66–70**
02031	01191	10820	5 06	05342	0303	0201	96212	00800	10101	09204	00101	01191	10007
		4	72201	01731	01023	20101	48112	02301	30100	89201	00301	00992	12002
02143	01191	15820	5 02	01972	0105	01	312	014					000
		3	65101	01092	04115	10101	52212	01058	50100	85101	58502	01981	05002
02231	01191	06920	3 01	00892	0101	0200	98112	014					000
		1	71101	01011	01161	10100	96109	01000	10101	59101	02901	01191	11005
02333	01 31	19748	5 01	00691	0185	0100	65112	02601	20101	19202	00301	00792	09003
		3	69601	01281	02369	60101	69109	01758	50100	89101	00101	01542	12010
02429	01191	06920	3 01	01291	0103	0101	08112	02500	10101	29105			001
		2	82101	00742	01094	10101	24212	05900	90101	53203	00901	01362	06002
02549	21221	04312	7 01	01081	0127	0101	09211	01200	10101	49101			001
		2	14101	01631	03303	10101	22106	00200	30201	98202	00302	01982	08003
02639	01193	16665	6 01	00891	0105	0101	14104	004					
								00000	00100	99101	00001	01391	09003
02739	01191	24660	5 01	01291	0105	0101	39112	019					000
		0	74101	00952	01052	10100	89212	02200	50100	73104	00501	00731	10002
02852	11161	16445	4 01	01791	0154	0101	69112	02800	30101	09101	00301	01091	09006
		2	42201	01391	01236	20101	79112	05000	50101	57101	00501	02891	12005
02969	01131	19669	8 01	01191	0104	0101	29112	03101	201	202	01201	01531	12006
		1	62201	02042	01265	20203	17212	12600	30100	99103	00301	01251	12004
03049	11191	10930	1 01	01851	0101	0101	59209	006					000
		0	74101	01291	01			00100	50100	99202	00101	01452	12002
03133	31131	11620	3 01	01171	0103	0101	09111	01200	10101	05204			001
		5	92101	01312	01041	20101	92112	04600	30101	39101	00301	01691	12009
03239	11191	16830	5 02	01632	0139	0100	19212	01400	10100	49207			001
		1	53101	01291	01175	10100	35207	00400	30100	88201	18601	00372	12007
03369	01191	21960	5 01	01191	0143	0100	79212	011					000
		2	32301	01411	01115	10101	59112	05500	30100	79201	53401	00991	11004
03423	31131	15820	1 02	00781	0354	0100	99211	01400	30101	39109	00101	01191	09002
		1	62201	01642	01369	60101	89112	06500	10101	49102	00301	00651	11002
03531	21141	11539	8 02	00741	0198	0401	40112	01800	10100	40201	00101	00992	12014
		1	62201	01891	01162	20101	39112	02855	50101	09103	00101	2	12009
03653	21191	18910	5 01	00882	0105	0100	99211	007					000
		2	82101	01182	01036	10101	09212	03301	80100	99208			001
03749	01191	10922	7 02	01982	0127	0302	97212	002					000
		2	73101	01412	02195	10101	45212	00900	50101	49204	05101	01092	10003
03859	11141	16452	7 01	00981	0118	0100	83212	021					000
		2	61201	01311	01162	20101	54212	03400	50101	45101	00501	01751	12004
03943	11291	15552	7 01	01081	0103	0100	88112	01300	70100	39105			001
		5	43201	01791	01115	10101	00212	01300	30100	83201	00301	01162	10003

(Continued)

Table 3 *Continued*

Record Columns													
01–05	**06–10**	**11–15**	**16–20**	**21–25**	**26–30**	**31–35**	**36–40**	**41–45**	**46–50**	**51–55**	**56–60**	**61–65**	**66–70**
04039	02232	04630	1 01	01591	0101	0100	49112	01100	10101	49112			001
		3	03101	01081	01303	10101	17112	00900	30101	49101	02701	00891	11014
04139	31131	16530	1 01	01292	0203	0201	98212	02601	201	202			001
		3	65101	01671	02203	20101	99112	02406	50100	95101	00501	01371	02002
04239	31191	17522	7 02	02181	0127	0201	48212	02400	10502	00203			001
		7	23202	04301	02			00100	30403	72203	57901	00931	08006
04352	01141	17250	1 03	00992	0101	0301	05112	021					000
		6	22201	02162	01282	10101	43212	04204	10100	99201	04101	01182	12013
04449	01291	15960	5 01	01091	0118	0101	29112	01100	30101	19101	00101	01581	11002
		1	73201	01091	01396	20101	79109	02006	50100	79101	00101	01591	12008
04539	12191	0674						00001	201	211			001
		0	52101	00392	01074	10100	67212	03256	50100	99204	00101	01172	11003
04639	01191	17940	3 01	00642	0127	0100	88212	03700	10101	49110			001
		0	74101	01191	01772	20101	77212	02600	30101	29202	18602	04332	12004
04729	01191	11830	3 01	01231	0105	0201	98212	02600	10101	19102			001
		1	83301	01721	02232	30102	66111	01556	90100	94111			001
04851	11191	17930	8 03	02372	0308	0602	62203	003					000
		0	74101	01491	05074	10101	59112	00803	70101	40109	00301	02822	09002
04939	04111	11240	4 01	00892	0118	0101	19212	02500	30100	69205	01201	00992	11002
		2	42201	01652	01772	20101	49212	03603	00100	87101	03001	01191	10006
05043	12132	16722	7 02	01582	0118	0201	71210	00901	20100	59201	01201	00442	09006
		7	03101	00792	01753	20101	59212	04300	30202	18201	00301	01791	12009
05153	31191	21253	9 01	00092	1239	0100	19212	00700	10100	95207			001
								00000	10100	89201	00101	01292	10004
05251	21113	18359	8 10	04902	0321	0100	79212	004					000
		9	49101	01591	03929	10100	50112	01006	20101	59104	06202	02982	12004
05333	02221	15555	4 01	01051	0143	0100	89212	02500	10101	29103			001
		0	52101	01131	01023	20101	79112	00702	50100	79103	00101	01191	12006
05449	21191	15730	3 02	02101	0103	0100	98212	012					000
		0	36101	01481	01868	20101	79112	00500	30101	27102	18601	01991	12012
05559	21141	13444	3 02	01981	0118	0201	98111	018					000
		2	32302	02781	01133	10101	80112	02000	50303	57203	00501	01011	12005
05651	31191	21850	5 01	00971	0105	0202	34112	022					000
		0	74101	01771	01074	10101	79112	03900	50101	49101	00501	01791	12009
05733	01111	11000	1 01	00391	0205	0100	49210	01100	10101	19102	00301	00692	07002
		1	53101	00792	03203	20101	74210	00700	30100	89103			001
05841	31311	17351	2 01	00991	0139	0100	88112	02200	10100	75201	58501	01031	07005
		0	52101	01581	01052	10101	55112	01200	30101	57101	00301	01292	11007
05939	31131	15730	5 01	00882	0103	0201	76209	00606	60201	46203	06602	01262	03002
		3	66201	01602	01152	10101	22112	04300	30100	79202	00301	00991	12009

(Continued)

Table 3 *Continued*

Record Columns													
01–05	**06–10**	**11–15**	**16–20**	**21–25**	**26–30**	**31–35**	**36–40**	**41–45**	**46–50**	**51–55**	**56–60**	**61–65**	**66–70**
06033	01191	16840	5 04	04362	0105	0303	27212	01601	20101	39106	00101	01292	11002
		1	15101	01391	01345	10101	82112	00804	10100	98202	00301	02391	12004
06139	01141	13330	1 01	01692	0201	0101	69211	01458	50100	69104			001
		6	82201	01532	01203	20102	23112	02700	30101	39201	00301	01481	10003
06259	03211	13368	5 02	00621	0185	0200	62112	027					000
		3	91201	01991	01592	10101	14212	02958	50100	79101	00701	01891	12013
06339	11191	13620	3 01	01191	0103	0201	94212	02300	10101	18203	00101	01312	11010
		1	53101	01122	01303	10100	54212	08900	30100	99201	00501	01482	12011
06459	11191	11931	8 02	02052	0105	0100	99211	00800	10101	28207			001
		3	03101	01291	02074	10101	27109	00500	30101	19103	00301	01291	08004
06539	13191	16925	6 01	01211	0108	0101	33112	013					000
		0	52101	01351	01365	10102	29112	02500	30102	72101	00601	02331	12019
06629	11191	06826	3 01	00591	0198	0100	59112	017					000
		1	08101	00651	01999	10100	94212	03806	50100	79101	02702	01382	12016
06729	11191	06940	3 01	00991	0105	0101	93112	01809	90100	24212			001
		0	66101	01991	03303	10100	52205	00300	10101	59103	50401	00991	07002
06829	11191	11002	7 01	01542	0227	0100	99211	00900	30101	59104			001
								00000	10101	49101	00301	01292	09007
06931	01191	17831	8 04	04362	0218	0303	57211	022					000
		2	05201	01451	01301	50101	49208	00800	50202	38208	00501	01392	11004
07049	03341	13550	1 02	00891	0185	0301	00112	05900	10101	19103	01201	00991	12011
		0	52101	01342	01622	20201	00212	07501	80100	99201	59201	01192	12007
07163	01131	23762	7 01	00892	0154	0100	69212	009					000
		1	23301	01891	01703	10101	39212	04700	50101	34201	00101	01492	12007
07239	03141	13233	9 01	00552	0130	0100	98112	01000	30100	99205			001
		3	32201	01841	04753	20102	09110	00700	31211	88201	02501	00992	12006
07369	01111	17452	7 02	01981	0327	01	312	010					000
		0	81301	01781	01052	10101	69212	10350	90100	95101	00301	01581	12006
07439	04191	08843	0 01	00492	0205	0201	76212	010					000
		9	08101	01101	02074	10100	93212	01700	30101	12201	03201	00692	12002
07539	01191	15830	5 10	06902	1105	0602	54212	00400	10101	26107			001
		0	36101	01751	06273	10101	58107	00400	30101	08112			001
07639	01191	15840	5 01	01151	0221	0100	79206	01000	30100	99202	00301	00742	11006
		2	32301	01402	01125	10101	48212	08100	30101	39101	00301	01542	07002

(Continued)

Table 3 *Continued*

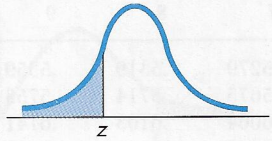
Record Columns													
01–05	06–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55	56–60	61–65	66–70
07739	31311	06135	4 01	00991	0154	0101	87112	02000	10101	49109			001
		2	26101	01211	01153	10101	82212	03306	50100	88101	06501	00991	12008
07829	14121	10326	3 01	00591	0163	0100	59112	01900	30101	29102	00301	01891	11008
		1	53101	02091	01052	10100	99112	03206	50100	69201	02701	00692	12010
07949	11121	17342	7 01	01642	0105	0101	66212	032					000
		1	33101	01391	01203	20204	78112	05204	10100	89201	04101	00992	10009
08031	31141	15240	5 02	01512	0143	0201	78211	00800	10100	77202	00101	00692	12005
		2	14101	00942	01396	20100	75212	05700	30201	58201	00301	00991	12007
08139	31211	13345	4 01	00892	0181	0200	78212	021					000
		7	03101	00792	01322	30102	49112	01600	60101	24201	02201	01492	12011
08249	11191	06944	3 01	00392	0443	0100	85210	002					000
								00000	10101	64210			001
08333	31291	11818	5 01	00791	0185	0100	79111	012					000
		6	22201	01291	01355	10101	69110	00300	50100	99102	02901	01291	09003
08431	01191	10830	5 01	00642	0139	0100	89212	01200	10100	68209			001
		0	52101	01171	01066	10100	92212	042					000
08551	31191	08912	7 02	01981	0218	0101	50211	01200	10101	28107	00101	01281	10003
		1	25101	01591	01125	10101	68111	04300	10101	12109	00301	01611	10002
08649	01191	18841	0 02	01762	0139	0201	98211	010					000
		2	26101	01092	01345	10101	94212	01702	90101	09202	00302	03132	10006
08739	*01191*	*13930*	*5 01*	*01072*	*0103*	*0100*	*69212*	*012*					*000*
		5	43201	01502	01622	20101	39109	01300	30101	19201	00301	01651	11007
08851	11111	08361	8 01	00992	0185	0100	68112	00800	10100	79205			001
		3	91201	01652	01592	10101	98112	02550	40101	67112			001
08949	04191	11930	3 02	01562	0203	0201	98210	008					000
		1	83301	01071	01183	30101	47212	00700	10100	37202	10001	00881	09004
09049	01191	10920	5 01	00792	0112	0100	49212	013					000
		7	03101	00792	01074	10101	83112	01104	10101	19103	00501	01751	12004
09119	11211	15630	5 02	01932	0212	0201	30212	013					000
		6	83201	00992	01415	20100	99212	04716	70201	72101	00101	01691	12006
09239	34111	13430	3 01	00892	1221	0100	79212	00600	10101	31104	56901	00891	11003
		3	03101	01011	01303	10100	99112	01900	30101	59101	54001	01351	11003
09323	11211	11130	5 01	01751	0114	0101	53111	010					000
		0	52101	01491	01543	20102	21112	03700	30100	83101	00301	02291	12015
09463	11131	19765	4 02	01981	0139	201	98112	022					000
		0	62201	01741	01342	20101	95112	03300	10101	70102	00101	01851	11004

(Continued)

Table 3 *Concluded*

Record Columns													
01–05	06–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55	56–60	61–65	66–70
09529	11191	10821	8 02	02082	0118	0101	59212	004					000
		3	66201	01092	01763	10101	79112	02000	50101	13202	05101	01591	12009
09629	01191	04009	8 02	01581	0339	0100	99112	00700	30101	29102	00101	01141	05006
		1	23301	01591	04123	30101	69110	00500	50101	89112			001
09729	31191	06912	1 02	01582	0256	1209	96204	004					000
		0	52101	01351	01226	10101	59112	01455	00100	99202	55001	00991	11007
09853	02211	11000	5 01	00892	0405	0100	88211	012					000
		2	96101	01552	03282	10100	90211	01000	30100	77202	02901	01471	06003
09959	11241	15350	1 02	00841	0139	0100	49212	023					000
		3	91201	00832	01753	20101	45212	03600	50102	49201	00501	01791	10010
10049	11141	11320	3 02	01942	0103	0201	94211	00900	10100	89208	00101	01172	08002
		1	53101	01022	01052	10101	59212	03800	10101	49202	00302	02482	11014

Appendix B *Areas Under the Standard Normal Curve (Areas to the Left)*



z	0	1	2	3	4	5	6	7	8	9
−3.0*	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
− .9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
− .8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
− .7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
− .6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
− .5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
− .4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
− .3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
− .2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
− .1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
− .0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

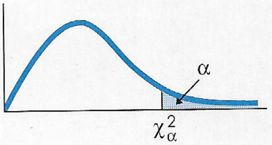
(Continued)

Appendix B *Concluded*

z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.9621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.9807	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0†	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Source: Neil Weiss, *Introductory Statistics,* © 1982, Addison-Wesley, Reading, Massachusetts. Pp. 576–577. Reprinted with permission.
*For $z \le -4$ the areas are 0 to four decimal places.
†For $z \ge 4$ the areas are 1 to four decimal places.

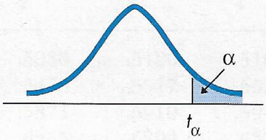
Appendix C *Chi-square Distribution (Values of χ^2_α)*



d.f. \ α	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.00	.00	.00	.00	.02	2.71	3.84	5.02	6.63	7.88
2	.01	.02	.05	.10	.21	4.61	5.99	7.38	9.21	10.60
3	.07	.11	.22	.35	.58	6.25	7.81	9.35	11.34	12.84
4	.21	.30	.48	.71	1.06	7.78	9.49	11.14	13.28	14.86
5	.41	.55	.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.68	.87	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	35.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2
500	422.3	429.4	439.9	449.1	459.9	540.9	553.1	563.9	576.5	585.2
1000	888.6	898.8	914.3	927.6	943.1	1058	1075	1090	1107	1119

Source: Neil Weiss, *Introductory Statistics,* © 1982, Addison-Wesley, Reading Massachusetts. P. 579. Reprinted with permission.

Appendix D *Student's t-Distribution (Values of t_α)*



d.f.	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$
1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.42	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.79
26	1.32	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
∞	1.28	1.64	1.96	2.33	2.58

Source: Neil Weiss, *Introductory Statistics,* © 1982, Addison-Wesley, Reading, Massachusetts. P. 578. Reprinted with permission.

Note: The last row of the table (d.f. = ∞) gives values for z_α. For example, the table shows that $z_{.10} = 1.28$ and $z_{.05} = 1.64$

Appendix E *Critical Values of the Sample Correlation Coefficient r*

α d.f.	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6	.622	.707	.789	.834
7	.582	.666	.750	.798
8	.549	.632	.716	.765
9	.521	.602	.685	.735
10	.497	.576	.658	.708
11	.476	.553	.634	.684
12	.458	.532	.612	.661
13	.441	.514	.592	.641
14	.426	.497	.574	.623
15	.412	.482	.558	.606
16	.400	.468	.543	.590
17	.389	.456	.528	.575
18	.378	.444	.516	.561
19	.369	.433	.503	.549
20	.360	.423	.492	.537
21	.352	.413	.482	.526
22	.344	.404	.472	.515
23	.337	.396	.462	.505
24	.330	.388	.453	.496
25	.323	.381	.445	.487
26	.317	.374	.437	.479
27	.311	.367	.430	.471
28	.306	.361	.423	.463
29	.301	.355	.416	.456
30	.296	.349	.409	.449
40	.257	.304	.358	.393
50	.231	.273	.322	.354
60	.211	.250	.295	.325
70	.195	.232	.274	.302
80	.183	.217	.257	.283
90	.173	.205	.242	.267
100	.164	.195	.230	.254

Source: Neil Weiss, *Introductory Statistics,* © 1982. Addison-Wesley, Reading, Massachusetts. P. 586. Reprinted with permission.

Note: A value given in the table is the *right-hand* critical value for a *two-tailed* test at the significance level indicated. The left-hand critical value is just the negative of the right-hand critical value.

Appendix F *Values of* $F_{.05}$

d.f. for Denominator	d.f. for Numerator								
	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

(Continued)

Appendix F *Concluded*

d.f. for Numerator									
10	12	15	20	24	30	40	60	120	∞
241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
3.64	3.57	3.51	3.41	3.41	3.38	3.34	3.30	3.27	3.23
3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Source: Neil Weiss, *Introductory Statistics,* © 1982, Addison-Wesley, Reading, Massachusetts. Pp. 584–585. Reprinted with permission.

Glossary

Achieved Communality A term used in factor analysis; it represents the proportion of variance in an original variable accounted for by all the extracted factors. Each original variable will have an achieved-communality value in the factor analysis output.

Acquiescence Bias (Yea-Saying) A bias stemming from the tendency on the part of certain respondents to agree with whatever side is presented by one-sided questions.

Ad Hoc Research Project A discrete, situation-specific project that is initiated and completed in response to a particular question, or set of related questions, raised by a decision maker.

Analysis of Covariance (ANCOVA) An advanced analysis of variance procedure in which the effects of one or more metric-scaled extraneous variables (called covariates) are removed from the dependent-variable data before one conducts the ANOVA.

Analysis of Variance (ANOVA) A technique for detecting relationships between a metric-scaled dependent variable and one or more categorical (nominal or ordinal) independent variables. ANOVA procedures are often used to analyze data from experimental studies.

Anchor Label A label used to define an extremity of a measurement scale.

Area Sampling A type of cluster-sampling procedure in which clusters are formed on the basis of the geographic location of the population units; that is, the clusters formed are subareas such as census tracts or counties.

Attitude A person's underlying mental state capable of influencing his or her choice of actions and maintaining consistency across those actions.

Attribute A personal or demographic characteristic of study respondents.

Audimeter An electronic device used by the A. C. Nielsen Company to monitor television-viewing behavior—what channels are watched, when, and for how long.

Balanced Scale A scale that has an equal number of positive/favorable and negative/unfavorable response choices.

Bar Chart A chart consisting of a series of bars (of equal thickness) whose heights (or lengths, if the bars are drawn horizontally) represent values of items.

Bayesian Analysis (Decision Theory Approach) A mathematical procedure for estimating the monetary value of a marketing research project.

Bayes's Rule A principle used in deriving posterior probabilities from prior and conditional probabilities; it states that for any research outcome R_i and market state M_j,

$$P(M_j/R_i) = \frac{P(R_i/M_j) \times P(M_j)}{P(R_i)}.$$

Behavioral Variable A variable that captures an activity engaged in by study respondents.

Belief A knowledge state of study respondents reflecting what they consider to be true.

Between Sum of Squares The portion of the total sum of squares that occurs due to differences between the treatment levels.

Bivariate Analysis A special case of multivariate analysis involving just two variables.

Blocking Factor A relevant characteristic of study units that serves as the basis for forming matched sets of units (called blocks) in certain experimental studies.

Cartoon (Balloon) Test A nonstructured, disguised form of questioning in which respondents are shown one or more cartoon pictures with empty "balloons" and are asked to fill the balloons with words reflecting thoughts or verbal statements of the characters involved.

Case (Observation) A sample unit for which data are available within a data set.

Case Study An in-depth examination, using a fairly flexible format, of a unit of interest such as a customer, store, salesperson, firm, or market area.

Census Study A study involving all units in a population.

Central Limit Theorem A theorem stating that for a sufficiently large sample size (in practice, a sample size of 30 or more), the sampling distribution curve for sample means associated with a sampling procedure will be centered on the population parameter value and will have all the properties of a normal probability distribution.

Central Tendency A number depicting the "middle" position in a given range or distribution of numbers.

Chi-square Contingency Test A nonparametric test for determining whether there is a statistically significant relationship between two categorical (nominal or ordinal) variables that have been cross-tabulated.

Chi-square Goodness-of-Fit Test A nonparametric test to check whether the observed distribution of sample data on a nominal-scaled variable is consistent with an expected distribution, that is, a distribution that would be expected if the sample had come from a prespecified population.

Classification Data Data generated by questions dealing with respondents' personal or demographic characteristics.

Client-Participant Group A focus group interview in which client personnel are introduced to respondents and made part of the group.

Cluster Analysis A technique for segmenting objects (e.g., customers, market areas, products) into homogeneous groups, given data for the objects on a variety of characteristics.

Cluster Sampling A probability sampling procedure in which clusters of population units are selected at random and then all or some of the units in the chosen clusters are studied.

Code of Ethics The standards governing the conduct of members of a profession.

Coding A term that broadly refers to the set of all tasks associated with transforming edited questionnaire responses into a form that is ready for analysis.

Coefficient of Determination (R^2) A global measure (varying between 0 and 1) of how good a regression equation is. It represents the proportion of variation in the dependent variable accounted for, or explained by, the independent variable(s) in the equation.

Comparative Rating Scale A scale that seeks respondents' feelings about an issue/object by explicitly asking them to compare the issue/object with a specified frame of reference.

Completely Nonstructured Question A question that is not necessarily presented in exactly the same wording to every respondent and does not have fixed responses.

Completely Randomized ANOVA An ANOVA procedure in which study units are randomly assigned to the treatments without explicitly matching the units across the treatments.

Completely Randomized Design An experimental study design in which (1) the study units are randomly divided into as many groups as there are experimental treatments and (2) the treatments are randomly assigned to the groups.

Completely Structured Question A question that is presented verbatim to every respondent and has fixed response categories.

Complex Question A question that is difficult to answer because it contains terms unfamiliar to respondents and/or requires considerable respondent effort.

Computer-Assisted Telephone Interviewing (CATI) A computerized telephone-interviewing system.

Computerized Interviewing A data collection procedure in which the questionnaire appears on a CRT terminal and the responses are directly entered into computer memory.

Computerized Search A computer-aided approach for quickly identifying potential sources of secondary data by simply specifying certain key words relating to the topic area on which data are desired.

Conclusive Research Research having clearly defined objectives and data requirements and capable of suggesting a specific course of action to be taken by decision makers.

Conditional Probability of a Research Outcome The probability that the research outcome would be observed if a particular market state or condition were to exist.

Confidence Interval A range of values (estimated from sample data) within which one can claim, with a prespecified degree of confidence, that a population parameter falls.

Confidence Level The minimum probability of not committing a Type I error when one is conducting a hypothesis test. The confidence level is 1 minus the specified significance level.

Confusion Matrix A table set up to evaluate the predictive accuracy of a discriminant function. It shows the degree of correspondence, or lack thereof, between the actual groupings of the sample units and the predicted groupings obtained by classifying the same units through the discriminant function.

Conjoint Analysis A technique for deriving the utility values that respondents presumably attach to various attributes of an object on the basis of respondents' overall preferences for different bundles of attributes or descriptive profiles of the object.

Consensor An electronic device used in certain focus groups. It consists of a set of hand-held terminals through which participants can express the nature and strength of their feelings about any issue raised by the moderator. The feelings of the group are instantly tabulated and displayed as a histogram on a video screen.

Constant-Sum Scale A scale that asks respondents to allocate a fixed number of points to various response categories in proportion to how the respondents feel about those categories (e.g., how important each category is or how much they like each category).

Constitutive Definition (Conceptual Definition) A definition that describes a construct in terms of other constructs.

Construct Validity A scale evaluation criterion that relates to the following question. "What is the nature of the underlying variable or construct measured by the scale?"

Content Validity (Face Validity) A scale evaluation criterion representing the extent to which the content of a measurement scale appears to tap all relevant facets of the construct it is attempting to measure.

Contingency Coefficient An index that measures the degree of association between two variables in a contingency table when the null hypothesis of independence between the variables is rejected.

Contingency Table A cross-tabulation set up for the purpose of conducting a chi-square contingency test.

Contrived Observation A data collection approach in which events and/or behaviors are observed in an environment artificially set up by the researcher.

Control Group Units in experiments that do not get exposed to the experimental treatment.

Controlling (Control Function) Getting feedback from the marketplace and taking needed corrective action.

Convenience Sampling A nonprobability-sampling procedure in which a researcher's convenience forms the basis for selecting a sample of units.

Convergent Validity A form of construct validity that represents the extent of association between a measured construct and measures of other constructs with which the measured construct is expected to be related on theoretical grounds.

Covariate An extraneous variable whose effects on the dependent variable are removed in conducting an analysis of covariance.

Cover Letter A letter to accompany a mail questionnaire in order to gain the cooperation of respondents.

Cramer's *V*-Statistic An index (varying between 0 and 1) that measures the degree of association between two variables in a contingency table when the null hypothesis of independence between the variables is rejected. This index is a somewhat more useful measure than the contingency coefficient.

Cross-Sectional Sample A group of units selected specifically and solely for data collection at a single period of time.

Cross-Sectional Study A one-time study involving data collection at a single period of time.

Cross-Tabulation A process involving simultaneous tabulation of data on two or more variables.

Crowded One-on-One Interview A variation of the depth interview, conducted by a professional interviewer in the presence of up to three client personnel, who themselves can probe the respondent toward the end of the interview.

Data Accuracy The extent to which the collected data are error-free and trustworthy.

Data Error A form of nonsampling error that represents any systematic bias occurring during the process of data collection, analysis, or interpretation.

Data Record A series of spaces in a data storage medium capable of accommodating a prespecified number of numerical digits.

Data Set (Data File) An organized collection of data records in which one or more records are set aside for storing data obtained from each unit participating in a study.

Decision Making Under Certainty Choosing a course of action after observing the results of perfect, or error-free, marketing research conducted to reduce uncertainty in a situation.

Decision Making Under Uncertainty Choosing a course of action without conducting any marketing research to reduce uncertainty in a situation.

Decision Rule A guideline that specifies the sample evidence necessary to reject the null hypothesis.

Decision Tree A diagram that lays out, in logical sequence, various available decision options, intervening events, and/or market states that can influence the monetary consequences of choosing each option.

Dependence Technique A multivariate analysis procedure in which one variable is designated as the dependent variable and the rest are treated as independent variables (e.g., regression analysis, ANOVA, discriminant analysis).

Dependent Variable (Criterion Variable) A variable in multivariate analysis that is presumed to be influenced by one or more independent variables. It is conventionally plotted along the *Y*-axis in a scatter diagram.

Depth Interview (In-Depth Interview) A form of nonstructured, nondisguised questioning in which respondents are encouraged to provide as much information, in as unrestricted a fashion, as possible.

Descriptive Analysis Preliminary data analysis (e.g., computing measures of central tendency and dispersion) that helps summarize the general nature of variables included in a study and the interrelationships among them.

Descriptive Research A form of conclusive research that is intended to generate data describing the composition and characteristics of relevant groups of units such as customers, salespeople, organizations, or market areas.

Dichotomous Question A structured question that offers just two answer choices.

Directional Hypothesis A null hypothesis that involves an inequality.

Direct Observation Observation that involves looking at a phenomenon rather than the phenomenon's consequences.

Discriminant Analysis A technique for generating a linear combination of independent variables (called a discriminant function) in such a way that values on it (called discriminant scores) are similar for sample units within prespecified subgroups and are maximally separated across the subgroups. The prespecified subgroup categories make up the categorical dependent variable in discriminant analysis.

Discriminant Axis A line that is perpendicular to the family of lines representing the discriminant function.

Discriminant Function A mathematical expression representing the linear combination of independent variables employed in a discriminant analysis.

Discriminant Score The discriminant-function value obtained for a unit by inserting the unit's independent-variable values into the discriminant function.

Discriminant Validity A form of construct validity that represents the extent to which a measured construct is *not* associated with measures of other constructs with which the measured construct is not expected to be related on theoretical grounds.

Discriminant Weight (Discriminant Coefficient) The weight (coefficient) corresponding to an independent variable in the discriminant function.

Disguised Observation Observation in which respondents are unaware they are being observed.

Disguised Question An indirect question whose true purpose is not obvious to respondents.

Disproportionate Stratified Random Sampling A stratified-random-sampling procedure in which the number of units chosen from each stratum is based on the degree of variability within it. The more heterogeneous a stratum, the larger is the sample chosen from it.

Double-Barreled Question A question that raises several separate issues but only provides one set of responses.

Dual-Moderator Group A type of focus group in which there are two moderators, one responsible for ensuring the smooth flow of the session and the other responsible for ensuring that key issues get discussed.

Dueling-Moderator Group A variation of the dual-moderator focus group. The two moderators in a dueling-moderator group deliberately take opposite positions on the discussion topic so as to stimulate a lively and potentially revealing exchange of views among the participants.

Dummy Variable A variable that takes on a value of either 1 or 0. Categorical variables that normally would not be included in certain analysis procedures (e.g., regression analysis) can be included in such procedures after converting the variables into a set of dummy variables. A categorical variable with k categories will require $k-1$ dummy variables to represent it.

Editing The process of examining completed data-collection forms and taking whatever corrective action is needed to ensure that the data are of high quality.

Eigenvalue A term used in factor analysis; it represents the total standardized variance in the original data set accounted for by a factor. Each extracted factor will have an eigenvalue in the factor analysis output.

End Piling A phenomenon wherein most responses fall into just a few categories at one end of a measurement scale.

Error Mean Square (Residual Mean Square) Error (residual)-sum-of-squares value divided by its degrees of freedom, $n-k-1$.

Error Sum of Squares (Residual Sum of Squares) The portion of the total sum of squares that is left unexplained by a regression equation.

Expected Value of Decision Making with Imperfect Information The expected payoff to a decision maker who commissions a less-than-perfect research project, observes the research outcome, and chooses the best course of action corresponding to that outcome.

Expected Value of Imperfect Information The value obtained by subtracting the expected value under uncertainty from the expected value of decision making with imperfect information; this value represents the monetary worth of a proposed research project that is less than perfect.

Expected Value of Perfect Information The value obtained by subtracting the expected value under uncertainty from the expected value under certainty; this value represents the maximum amount a decision maker should be willing to pay for any research aimed at reducing uncertainty in a situation.

Expected Value Under Certainty The expected payoff to a decision maker who commissions a perfect research project, observes the research outcome, and chooses the best course of action corresponding to that outcome.

Expected Value Under Uncertainty The maximum expected payoff to a decision maker who chooses a course of action under uncertainty.

Experiment A procedure in which one (or sometimes more than one) causal variable is systematically manipulated and data on the effect variable are gathered, while controlling for other variables that may influence the effect variable.

Experimental Group Units in experiments that get exposed to the experimental treatment.

Experimental Research (Causal Research) A form of conclusive research that is intended to generate the type of evidence necessary for confidently making causal inferences about relationships between variables.

Expert System A computer-based, artificial intelligence system that performs tasks normally associated with human intelligence.

Exploratory Research Research intended to develop initial hunches or insights and to provide direction for any further research needed.

External Providers of Marketing Research Research suppliers such as commercial marketing research firms and consultants.

External Secondary Data Data available from a source outside an organization for which a project is being conducted.

External Source A source outside the organization.

External Validity The extent to which results observed in an experiment are likely to hold beyond the experimental setting.

Eye-Tracking Equipment A device used in some mechanical-observation studies to ascertain which sections of a stimulus (e.g., an ad, product packaging, special display) respondents look at and how much time they spend looking at those sections.

Factor Analysis A technique that analyzes data on a relatively large set of variables and produces a smaller set of factors, which are linear combinations of the original variables, so that the set of factors captures as much information as possible from the original data set.

Factorial ANOVA An ANOVA procedure in which *two or more* categorical independent (treatment) variables are involved.

Factorial Design An experimental-study design in which two or more independent (causal) variables are manipulated simultaneously.

Factor Loading A term used in factor analysis; it represents the Pearson correlation coefficient between an original variable and a factor.

Factor-Loading Matrix A matrix containing the pairwise correlations between a set of original variables and the set of factors derived through factor analysis.

Factor Score A term used in factor analysis; it represents the value for a sample unit on a given factor and is obtained by substituting the sample unit's values on the original variables into the linear combination representing the factor.

Field Experiment A research study conducted in a natural setting in which one or more independent variables are manipulated by the experimenter under as carefully controlled conditions as the situation will permit.

Field Service Firm A data gathering firm that caters to research suppliers as well as in-house marketing research departments.

Filter Question A question meant to qualify respondents for a subsequent question or to ensure that the question is within their realm of experience.

Final Edit (Office Edit) A thorough examination of all completed data collection forms done in the office; it involves verifying response consistency and accuracy, making necessary corrections, and deciding whether some or all parts should be discarded.

Focus Group Interview A setting in which respondents (typically about 8 to 12) discuss a given topic in a fairly informal fashion in the presence of a well-trained, objective moderator.

Forced-Choice Scale A scale that does not give respondents the option of expressing a neutral opinion.

Four-Group, Six-Study Design A true experimental design in which two experimental groups and two control groups are formed by random assignment, one experimental group and one control group are premeasured, and all four groups are postmeasured.

Fourth-Generation Language A computer software package that allows even individuals with no training in computer programming to directly access and manipulate data bases.

Full-Profile Approach A data-gathering procedure used in conjoint analysis, wherein respondents rank-order their preferences for complete stimuli, each of which is a different combination of levels of *all* attributes.

Funnel Sequence An ordering of questions that begins with a very general question on a topic and gradually leads up to a narrowly focused question on the same topic.

Galvanic Skin Response (GSR) Meter A device for measuring changes in electrical resistance of the skin caused by changes in amount of perspiration.

Graphic Rating Scale A continuous scale, presented in the form of a straight line, along which a theoretically infinite number of response choices are possible.

Histogram A two-dimensional bar chart used for graphically displaying the absolute or relative frequency of occurrence of values pertaining to a variable. The variable values are plotted along the horizontal axis, and their frequencies of occurrence are plotted along the vertical axis.

History Effect A bias that occurs because of specific external events or occurrences during an experiment (e.g., changes in competitors' marketing mixes) that are likely to affect the dependent variable.

Hit Rate The porportion of units correctly classified by a discriminant function.

Human Observation A method of observation using human observers (as opposed to mechanical or electronic devices) to collect data.

Hybrid-Sampling Plan A sampling plan that incorporates aspects of both probability- and nonprobability-sampling methods.

Ideal Population The collection of units most pertinent for the study; it is the population to which generalizations from the study should be desired.

Implied Population The collection of units to which any generalization from the actual sample obtained in a study is restricted.

Incidence Rate In sampling, the proportion of the general population possessing a particular characteristic.

Independent Variable (Explanatory or Predictor Variable) A variable in multivariate analysis that is presumed to influence a dependent variable. It is conventionally plotted along the X-axis in a scatter diagram.

Indirect Observation Observation that involves looking at the results or consequences of a phenomenon rather than directly observing the phenomenon itself.

Inferential Analysis Data analysis that goes beyond descriptive analysis; it involves verifying specific statements or hypotheses about the population(s).

Information The product of analysis and interpretation of data.

Information Consultant An individual with special skills pertaining to effective secondary-data management.

In-House Marketing Research Research conducted by a firm's own marketing research department or employee(s) in charge of marketing research.

Instrument Variation Effect A bias in experimental studies that relates to differences between pretest and posttest measurements owing to changes in the instruments (questionnaires) and/or procedures used to measure the dependent variable.

Interaction Effect Effect in an experiment when the relative impact of different levels of one independent variable depends on the levels of another independent variable.

Interdependence Technique A multivariate analysis procedure in which there are no dependent- and independent-variable designations; all variables are treated equally in a search for underlying patterns of relationships (e.g., factor analysis, cluster analysis).

Internal Secondary Data Data available from a source within an organization for which a project is being conducted.

Internal Source A source within the organization.

Internal Validity The extent to which results observed in an experiment are solely due to the experimental manipulation.

Interval Scale A set of numbers in which the differences (but not ratios) between numbers can be meaningfully interpreted.

Inverted-Funnel Sequence An ordering of questions that begins with very specific questions on a topic and moves on to more general questions on the same topic.

Itemized Rating Scale A scale that has a finite set of distinct response choices.

Judgment Sampling (Purposive Sampling) A nonprobability-sampling procedure in which a researcher exerts some effort in selecting a sample that he or she feels is most appropriate for a study.

Key-Informant Technique (Expert-Opinion Survey) Conducting exploratory research by seeking out and talking to individuals with expertise in areas related to the situation being investigated.

Kolmogorov-Smirnov Test A nonparametric test to check whether the observed distribution of sample data on an ordinal-scaled variable is consistent with an expected distribution. This test is similar to the chi-square goodness-of-fit test, except that it requires ordinal data.

Laboratory Experiment A research study conducted in a contrived setting in which the effect of all, or nearly all, possible influential independent variables not pertinent to the immediate problem is kept to a minimum.

Latin-Square ANOVA An ANOVA procedure in which *two* blocking factors, each having the same number of levels as the number of treatments, are employed simultaneously.

Latin-Square Design An experimental-study design in which two blocking factors are employed simultaneously. In this design the number of levels of each blocking factor must be the same as the number of treatments.

Leading Question (Loaded Question) A question that tends to steer respondents toward a certain answer, irrespective of what their true answers are.

Least Squares Approach An analysis procedure widely used in deriving a regression equation that is the best-fitting equation for a given set of data on dependent and independent variables.

Likert Scale A multiple-item scale consisting of a series of evaluative statements concerning an issue or object; responses to each statement are provided on a five-point, agree-disagree scale.

Line Chart A two-dimensional graph typically used to show movements in one or more items over time.

Longitudinal Study A repeated-measurement study involving data collection at several periods in time.

Mail Survey A questionnaire administration method in which there are no interviewers and no direct contact between the researcher and respondent.

Main Effect The distinct effect of an independent variable on the dependent variable in an ANOVA procedure, regardless of the levels of other independent variables.

Manipulation Check Formal measurement of the independent (causal) variable in marketing experiments.

Marketing Concept The philosophy of customer orientation urging firms to uncover customer needs first and then coordinate all their activities to satisfy those needs.

Marketing Decision Support System A marketing information system that permits managers to interact with it and request special types of data analyses or reports on an as-needed basis.

Marketing Information System A continuing and interacting structure of people, equipment, and procedures designed to gather, sort, analyze, evaluate, and distribute pertinent, timely, and accurate information to marketing decision makers.

Marketing Mix The set of product, promotion, price, and distribution (or place) decisions related to a firm's market offering.

Marketing Research A set of techniques and principles for systematically collecting, recording, analyzing, and interpreting data that can aid decision makers who are involved with marketing goods, services, or ideas.

Marketing Research Ethics Standards that govern the marketing research profession.

Marketing Response Function A mathematical model that represents the relationship between marketing input and output variables.

Market State The overall nature or disposition of the market.

Matching A procedure used in certain experimental studies to form groups of units in such a way that the group compositions are similar with respect to some specific characteristic(s).

Maturation Effect A bias that occurs because of the impact on the dependent variable of physiological or physical changes in units participating in an experiment.

Maximum Chance Criterion The proportion of correct classifications that would result if *all* sample units are classified into the larger of two prespecified groups.

Mean A measure of central tendency that represents the simple average of a set of numbered responses.

Mean Square A sum-of-squares value divided by its corresponding degrees of freedom.

Measurement The quantification of observations or responses by assigning numbers to them according to a given set of rules.

Measurement Level The extent of precision of survey data in terms of whether they form a nominal, ordinal, interval, or ratio scale.

Mechanical Observation Observation in which mechanical or electronic devices (as opposed to human observers) are used to collect data.

Median A measure of central tendency that represents the category in which the 50th percentile response falls when all responses are arranged from lowest to highest (or vice versa).

Metric Data Data having interval- or ratio-scale properties.

Missing-Value Category A response category used to code questions for which answers should have been obtained but were not for some reason.

Mode A measure of central tendency that represents the most frequently occurring response category.

Mortality Effect A bias that occurs when certain participating units drop out of an experiment, and, as a result, the set of units completing the experiment significantly differs from the original set of units.

Motivation Research Research intended to uncover inner feelings and motives that people are unable or unwilling to reveal when questioned directly.

Multicollinearity A condition wherein two or more independent variables in a multiple regression equation are highly correlated.

Multidimensional Scaling A technique for inferring the number and the nature of dimensions underlying respondent evaluations/perceptions on the basis of similarity and/or preference judgments provided by respondents about objects (e.g., companies, products, brands).

Multiple-Category Question A structured question that offers more than two fixed-response choices.

Multiple Correlation Coefficient A measure of the overall degree of association between dependent and independent variables in a regression equation; it is simply the square root of the coefficient of determination (R^2).

Multiple-Item Scale A scale that consists of two or more questions pertaining to the construct it is intended to measure.

Multiple Regression Analysis A technique that generates a mathematical function or equation linking a dependent variable with *two or more* independent variables.

Multistage Area Sampling A form of area sampling in which more than two sequential stages of random selection are involved in the process of choosing the sample.

Multivariate Analysis Any procedure in which the focus is on analyzing data on two or more variables simultaneously.

Natural Observation A data collection approach in which events and/or behaviors are observed as they occur naturally.

Nominal Scale A set of numbers in which the numbers serve solely as identification labels.

Noncomparative Rating Scale A scale that seeks respondents' feelings about an issue/object without providing any frame of reference with which to compare the issue/object.

Nondirectional Hypothesis A null hypothesis that involves a strict equality.

Nondisguised Observation Observation in which respondents are aware they are being observed.

Nondisguised Question A direct question whose true purpose is obvious to respondents.

Nonforced-Choice Scale A scale that gives respondents the option of expressing a neutral opinion.

Nonmetric Data Data having only nominal- or ordinal-scale properties.

Nonparametric Procedures Analysis techniques and hypothesis tests that are appropriate when the data are nonmetric (i.e., nominal or ordinal).

Nonprobability Sampling A subjective-sampling procedure in which the probability of selection for each population unit is unknown beforehand.

Nonrepresentative-Sample Bias A bias that occurs in experimental studies when the participating units are not representative of the larger body of units to which the results are to be generalized.

Nonresponse Error A bias that occurs when the final sample differs in a systematic way from the planned sample; that is, when respondents and nonrespondents differ significantly on one or more critical characteristics.

Nonsampling Error Any error in a research study other than sampling error (which arises purely because a sample, rather than the entire population, is studied).

Nonstructured Observation Observation in which everything of relevance is observed and recorded in an open-ended fashion.

Not-at-Home Problem A potential source of nonresponse error; it is serious when respondents who are not at home when the interviewer makes contact differ significantly from those who are at home.

Observation A primary-data collection approach in which respondents play little or no active role in terms of interacting with the data gatherer.

Omnibus Panel A panel that serves as a ready source of samples for a variety of cross-sectional studies over time.

One-Group, After-Only Design A pre-experimental design in which one group of units is exposed to a manipulation and then measured.

One-Group, Before-After Design A pre-experimental design in which one group of units is measured, exposed to a manipulation, and then measured again.

One-Sided Question A form of leading question that presents only one aspect of an issue for which respondents' reactions are being sought.

One-Stage Area Sampling A form of area sampling in which a few subareas are chosen at random and all the units in the chosen subareas are studied.

One-Tailed Test (Hypothesis Test) A hypothesis test in which values of the test statistic leading to rejection of the null hypothesis fall only in one tail of the sampling distribution curve.

One-Way Table A table showing the distribution of data pertaining to categories of a single variable.

Open-Ended Question A question that respondents are free to answer in their own words.

Operational Definition A definition that describes a construct in terms of how it is to be measured.

Ordinal Scale A set of numbers in which the magnitudes of the numbers represent no more than a rank ordering.

Original Source A term used in the context of secondary data; it refers to the source that actually collects the data.

Paired-Comparison Rating Scale A scale that seeks comparative evaluations of two issues/objects at a time.

Panel A fixed group of units recruited to provide measurements over a period of time.

Panel Conditioning A bias that occurs when multiple-survey participation by panel members, over a period of time, induces them to alter their natural or usual behavior.

Panel Design Experimental design that uses data gathered from consumer and other types of panels.

Parameter A summary measure based on data from all units in a given population.

Parametric Procedures Analysis techniques and hypothesis tests that are appropriate only when the data are metric (i.e., interval or ratio).

Payoff Table A table summarizing the net profits (or losses) under various possible combinations of decision alternatives and market states.

Pearson Correlation Coefficient (Pearson Product-Moment Correlation) An index (varying between -1 and $+1$) that measures the nature and the degree of association between two metric-scaled (interval or ratio) variables.

People Meter An electronic device used by the A. C. Nielsen Company to monitor television-viewing behavior—who is watching, in addition to what shows are being watched.

Periodicity Any cyclical pattern with respect to a relevant characteristic in a population list used in systematic sampling. The statistical efficiency of systematic sampling will be lowered when the periodicity coincides with the sampling interval, a multiple of the sampling interval, or an integral fraction of the sampling interval.

Personal Interview A questionnaire administration method in which the interviewer and respondent have face-to-face contact.

Pictogram A special type of bar chart that uses pictures of the items instead of the customary rectangular bars.

Pie Chart A circle divided into several slices whose areas are proportionate to the quantities to be represented on the chart.

Plus-One Dialing A method of reducing sampling frame error in telephone surveys; it involves adding a fixed integer (usually 1) to each telephone number chosen from a phone book so as to obtain a revised number to be dialed.

Point Estimate A dependent-variable predicted by using a regression equation.

Population (Universe) The entire body of units of interest to decision makers in a situation.

Posterior Probabilities A set of revised probabilities for the various market states or conditions, contingent upon a particular research outcome being observed.

Power The probability that a hypothesis-testing procedure will lead to rejection of a null hypothesis that is false. Power is 1 minus the probability of committing a Type II error, which is represented by the symbol β; that is, power $= 1 - \beta$.

Precision Level A measure of the degree of compactness of a confidence interval. The narrower the confidence interval, the higher the precision level is.

Prediction Error The difference between the actual value of a dependent variable and its estimated value obtained by using a regression equation.

Prediction Interval A confidence interval constructed around a point estimate (for the dependent variable) generated by a regression equation.

Predictive Validity A scale evaluation criterion representing how well a measured construct is able to predict some other variable or characteristic it is supposed to influence.

Pre-Experimental Design An experimental design that offers little or no control over the influence of extraneous factors and hence is not much better than a descriptive-research design when it comes to making causal inferences.

Preliminary Edit (Field Edit) A quick examination of completed data collection forms, usually on the same day they are filled out.

Pretesting An important stage in the questionnaire refinement process, prior to finalizing the questionnaire; it involves administering the questionnaire to a limited number of potential respondents and other knowledgeable individuals in order to identify and correct design flaws.

Pretesting Effect A bias in experimental studies that occurs when responses given during a later measurement are influenced by those given during a previous measurement, regardless of what happens between the measurements.

Pretest-Manipulation Interaction Bias A bias that arises when the premeasurement increases or decreases respondents' sensitivity to the experimental manipulation.

Primary Data Data collected specifically for a project.

Prior Probability The probability that a particular market state or condition exists, based purely on the subjective feelings of a decision maker.

Probability-Proportional-to-Size Area Sampling A form of two-stage area sampling in which the probabilities of selection for the subareas in the first stage are adjusted to be proportional to the number of second-stage units in each.

Probability Sampling An objective-sampling procedure in which the probability of selection is known in advance for each population unit.

Professional Respondent A person who has already participated in numerous focus group interviews.

Projective Technique A nonstructured, disguised form of questioning in which a fairly ambiguous stimulus is presented to respondents, who are then asked to react to or describe the stimulus.

Proportional Chance Criterion The proportion of correct classifications that would result if each unit in a sample is probabilistically classified into one of two prespecified groups solely on the basis of the relative sizes of the two groups.

Proportionate Stratified Random Sampling A stratified-random-sampling procedure in which the number of units chosen from each stratum is proportional to its size; the more the number of units in a stratum, the larger is the sample chosen from it.

Pseudo-Research Research requested by managers for reasons other than aiding their decision making.

Pupillometer A device for measuring the extent of pupil dilation in response to a visual stimulus.

Qualitative Research A form of exploratory research involving small samples and nonstructured data collection procedures.

Quantitative Research A form of conclusive research involving large representative samples and fairly structured data collection procedures.

Quasi-Experimental Design An experimental design that does not offer as much control as a true experimental design but usually provides more measurements and more information than a typical preexperimental design.

Questioning (Communication) A primary-data collection approach in which respondents play an active role owing to their interaction with the data gatherer.

Questionnaire An instrument used for eliciting and recording responses in many, but not all, research projects employing the questioning approach.

Question with Implicit Assumptions A question that does not provide or imply the same frame of reference to all respondents.

Quota Sampling A nonprobability-sampling procedure in which (1) the population is divided into cells on the basis of relevant control characteristics, (2) a quota of sample units is established for each cell, and (3) interviewers are asked to fill the quotas assigned to the various cells.

Random Assignment A procedure used in experimental studies to form groups of units in such a way that the group compositions can be considered to be equivalent in all respects. This term also refers to assigning experimental treatments randomly to the various groups.

Random-Digit Dialing A method of reducing sampling frame error in telephone surveys; it involves randomly dialing the last four digits of telephone numbers within any exchange specified by a three-digit prefix.

Randomized-Block ANOVA An ANOVA procedure in which study units are divided into homogeneous blocks on the basis of a blocking factor and the units in each block are randomly assigned to the treatments.

Randomized-Block Design An experimental study design in which (1) the study units are divided into blocks, wherein each block is homogeneous (i.e., has units that are matched on some relevant characteristic) and contains as many units as there are treatments, and (2) the units within each block are randomly assigned to the treatments.

Random Sampling An objective procedure for selecting a representative sample of units for a study.

Ratio Scale A set of numbers in which the ratios between numbers can be meaningfully interpreted.

Raw Variable A variable that is directly measured by the data in a questionnaire.

Reactive Bias A bias that occurs in experimental studies when participants exhibit abnormal or unusual behavior simply because they are participating in an experiment.

Regression Analysis A technique that generates a mathematical function or equation linking one dependent variable with one or more independent variables.

Regression Mean Square Regression-sum-of-squares value divided by its degrees of freedom, k.

Regression Sum of Squares The portion of the total sum of squares that is explained by a regression equation.

Reliability A criterion for evaluating measurement scales; it represents how consistent or stable the ratings generated by a scale are.

Research Continuum A line or dimension along which conclusive research projects can be arrayed, with "purely descriptive with no control" at one extreme and "purely experimental with strict control and manipulation" at the other.

Research Process An interrelated sequence of steps that make up a research project.

Research Proposal A document that briefly describes the following aspects of the contemplated project: purpose and scope, specific objectives, sample design, data collection procedures, data analysis plan, timetable, and estimated cost.

Research Subject (Respondent) An individual from and/or about whom data are gathered.

Research Supplier A commercial marketing research firm that provides information to a client firm.

Respondent-Moderator Group A focus group interview in which the moderator may ask selected participants to play the role of moderator to improve group dynamics and productivity.

Respondent-Refusal Problem A potential source of nonresponse error; it is serious when respondents who refuse to participate in a study differ significantly from those who agree to participate.

Response Category Sequence Rotation The process of varying from respondent to respondent the order in which response choices to a multiple-category question are presented.

Response Latency The speed with which a respondent provides an answer.

Response Rate The percentage of units in the planned sample for a study that actually participated in the study.

Sampling The process of selecting a fraction of the total number of units of interest to decision makers for the ultimate purpose of being able to draw general conclusions about the entire body of units.

Sampling Distribution A representation of sample statistic values—obtained from every conceivable sample of a certain size chosen from a population by using a specified sampling procedure—along with the relative frequency of occurrence of those statistic values.

Sampling Error A measure of the difference between a statistic value that is generated through a sampling procedure and the parameter value, which can only be determined through a census study.

Sampling Frame A list or collection of population units from which the sample for a study is chosen.

Sampling Frame Error A bias thast occurs when the population as implied by the sampling frame differs in a systematic fashion from the ideal population.

Sampling Interval The size of the gap between units chosen from a population list by using a systematic-sampling procedure.

Scatter Diagram A two-dimensional graph in which data on two metric-scaled (interval or ratio) variables are plotted.

Secondary Data Data collected for some purpose other than the research situation.

Secondary-Data Management Creating and operating a system for continuously monitoring various secondary-data sources and quickly retrieving needed data.

Secondhand Source A term used in the context of secondary data; it refers to a source that uses data collected by an original source to generate its own summaries, interpretations, and the like.

Selection Effect A bias that occurs when multiple groups participating in an experiment differ on characteristics that have a bearing on the dependent variable.

Semantic-Differential Scale A multiple-item scale consisting of a series of bipolar adjectival words or phrases pertaining to an issue or object; responses to each pair of opposite adjectives are provided on a seven-category scale separating the adjectives (with no intermediate numerical or verbal labels).

Sensitivity A criterion for evaluating measurement scales; it represents the extent to which ratings provided by a scale are able to discriminate between respondents who differ with respect to the construct being measured.

Sentence Completion Test A nonstructured, disguised form of questioning in which respondents are asked to finish a set of incomplete sentences. A typical sentence completion test is given to a group of respondents who are asked to finish the sentences in writing.

Shopper Study A study in which a professional interviewer plays the role of a customer and gathers relevant data (through questioning as well as observation) on issues of interest.

Shopping Mall Intercept Interview A form of personal interviewing in which respondents are intercepted in shopping malls and interviewed.

Significance Level Represented by the symbol α; it is the maximum probability of rejecting a true null hypothesis when one is conducting a hypothesis test; it is the maximum probability of committing a Type I error.

Simple Cluster Sampling (One-Stage Cluster Sampling) A probability-sampling procedure in which (1) the population is divided into subpopulations, or clusters, (2) a simple random sample of clusters is selected, and (3) all units in the selected clusters are studied.

Simple Random Sampling A probability-sampling procedure in which every possible sample of a certain size within a population has a known and *equal* probability of being chosen as the study sample.

Simple Regression Analysis A technique that generates a mathematical function or equation linking a dependent variable with *one* independent variable.

Simulated Test Market A hybrid experimental approach involving a laboratory setting that imitates real market conditions as much as possible.

Single-Item Scale A scale that consists of just one question pertaining to the construct it is intended to measure.

Single-Source Data Base Contains, in an integrated fashion, data for individual households on household characteristics, product purchases, and exposure to marketing stimuli.

Skewed Response Distribution A one-way table in which a large proportion of responses are piled up toward one end of the range of data obtained.

Slope The regression coefficient (*b*-value) corresponding to an independent variable in a regression equation; it represents the change in the predicted value of the dependent variable per one-unit change in the independent variable, assuming all other variables likely to influence the dependent variable remain the same.

Spearman Correlation Coefficient An index (varying between -1 and $+1$) that measures the nature and the degree of association between two ordinal variables.

Split-Ballot Technique A procedure wherein (1) the study sample is split into two equivalent halves and (2) a different version of the same questionnaire (e.g., different sequencing of questions) is administered to each half.

Split-Half Reliability An indicator of the degree of consistency across ratings produced by items within a multiple-item scale.

Standard Deviation A measure of dispersion of a set of numbered responses from their mean.

Standard Error An indicator of the extent of sampling error; it represents the standard deviation of sample statistic values that will be obtained through repeated sampling from a population by using the same procedure.

Standard Error of the Regression An indicator of the dependent-variable variance left unexplained by the independent variable(s) in the equation; it can be viewed as a standard deviation of the prediction errors made by the regression equation.

Standardized Regression Coefficient Value of the slope when data on all the variables in a regression equation are standardized (reduced to a mean of zero and a standard deviation of one).

Stapel Scale A multiple-item scale consisting of a series of words or phrases pertaining to an issue or object; responses to each item are provided on a 10-category scale containing numerical labels for the categories.

Statistic A summary measure based on data from a sample of units.

Statistical Design A term used to denote an experimental design that measures the relative effectiveness of multiple levels of one or more independent variables and requires somewhat complex data analysis procedures.

Statistical Efficiency A measure of how good a sampling procedure is in terms of the sampling error associated with it. One sampling procedure is statistically more efficient than another if for a given sample size it results in a smaller sampling error.

Stratified Random Sampling A probability-sampling procedure in which (1) the population is divided into homogeneous strata on the basis of some appropriate characteristic and (2) a random sample of units is chosen from each stratum to make up the total sample.

Stratum Chart A two-dimensional graph with time along the horizontal axis and containing a series of distinct strata, or layers, stacked in the vertical direction. Each stratum represents a particular item, and at any given time the width of the stratum indicates the relative magnitude of the item at that time.

Structured Observation Observation in which the data gathered pertain to discrete, clearly defined events or behaviors and are therefore recorded by merely making entries or checking off categories on an appropriate observation form.

Syndicated Data (Syndicated Services) Secondary data generated by research suppliers on a regular basis and sold to many clients.

Syndicated Source A firm that offers syndicated data or services.

Systematic Error Any error that is not random; that is, any error that results in a definite upward or downward bias in the estimates obtained from a study.

Systematic Sampling A probability-sampling procedure in which (1) a sampling interval *k* is determined as the ratio of population size to desired sample size, (2) one unit between the first and *k*th units in the population list is chosen randomly, and (3) every *k*th unit thereafter in the list is chosen.

Systems-Oriented Marketing Research Making available appropriate marketing information on a regular, integrated basis.

Telephone Survey A questionnaire administration method in which there is only voice contact between the interviewer and respondent.

Test Marketing A form of field experiment for assessing a market's reactions to a new product and its associated marketing mix.

Test-Retest Reliability An indicator of the stability of ratings produced by a scale over time.

Test Statistic A term used in the context of hypothesis testing; it represents a standard variable whose value is computed from sample data and compared with a critical value (obtained from an appropriate probability table) to decide whether or not to reject the null hypothesis.

Thematic Apperception Test (TAT) A nonstructured, disguised form of questioning in which respondents are shown a series of pictures, one at a time, and are asked to write a story about each.

Total Error The combination of sampling and nonsampling errors in a study.

Total Sum of Squares Quantity obtained by adding across all data points the squared deviation of each actual value of the dependent variable from its mean value.

Transformed Variable (Recoded Variable) A variable that is created by combining or modifying data on one or more raw variables.

Treatment Each specific level at which an independent variable is set in an ANOVA procedure.

True Experimental Design An experimental design that has built-in safeguards for controlling all, or almost all, threats to internal and external validity.

True Panel A panel participating in a longitudinal study involving measurement of the same sample units, with respect to the same variables, over time.

Two-Factors-at-a-Time Approach (Trade-Off Approach) A data-gathering procedure used in conjoint analysis, wherein respondents rank their preferences for various combinations of attribute levels, considering just a pair of attributes at a time.

Two-Group, After-Only Design A true experimental design in which an experimental group and a control group are formed by random assignment and measured after the experimental group's exposure to the manipulation.

Two-Group, Before-After Design A true experimental design in which an experimental group and a control group are formed by random assignment, measured before the experimental group's exposure to the manipulation, and then measured again after the exposure.

Two-Group, *Ex' Post Facto* Design A pre-experimental design in which a manipulation is introduced among a set of units and then the set is divided into an experimental group and a control group on the basis of postmeasurement results.

Two-Stage Area Sampling A form of area sampling that involves random selection at two stages: (1) in selecting a sample of subareas and (2) in selecting a sample of units from each chosen subarea.

Two-Tailed Test (Hypothesis Test) A hypothesis test in which values of the test statistic leading to rejection of the null hypothesis fall in both tails of the sampling distribution curve.

Two-Way Table A table showing the number of responses in each category of one variable falling into the categories of a second variable.

Type I Error A mistake that occurs in a hypothesis test when a true null hypothesis is rejected.

Type II Error A mistake that occurs in a hypothesis test when a false null hypothesis is not rejected.

Unbalanced Scale A scale in which a majority of the choices favor one side of the issue being probed.

Unbiased Estimate A sample statistic whose average value (over repeated samples) coincides with the population parameter it is intended to estimate.

Univariate Analysis Any procedure in which the focus is on analyzing data on just one variable.

Utility Value A number (obtained through conjoint analysis) representing a respondent's degree of preference for a certain level of a given attribute.

Validity A criterion for evaluating measurement scales; it represents the extent to which a scale is a true reflection of the underlying variable or construct it is attempting to measure.

Voice Pitch Analysis (VOPAN) A technique for assessing how much emotional commitment is attached to a respondent's verbal answers; it involves measuring the voice pitch of verbal responses during an interview and comparing it with the respondent's normal voice pitch during routine conversation about neutral topics.

Within Sum of Squares The portion of the total sum of squares that occurs solely due to variation within the treatment groups.

Word Association Test A nonstructured, disguised form of questioning in which respondents are asked to express the first word that comes to mind in response to each word on a list of words. In a typical word association test the words are read aloud, one at a time, to each respondent.

***Y*-Intercept** The constant a in a regression equation; it represents the predicted value of the dependent variable when all independent variables are set to zero.

Index

Aaker, David A., 244n.42, 764n.12
ABC News—Harris Survey, 473
Abrams, Jack, 431n.23
Abstracts, as sources of secondary data, 189–190, 203–205
Achieved communality, 764
Acquiescence bias, 382
Action Market Research Services, Inc., *case* 72–74
Adams, Arthur J., 499n.6
Ad hoc research project, 41, 43
Adler, L., 35n.1, 37n.3, 268n.35, 270n.36
Agnew, Joe, 59n.25
AIM service, 185–187, 188
Akaah, Ishmael, 776n.24
Akey, Denise, 188n.13
Albaum, Gerald, 440n.33
Aldendick, Robert W., 382n.15
Allison, Neil K., 327n.13
Alpha level. *See* Significance level
Alternative hypothesis, 645
American Airlines, 50
American Chicle Youth Poll, 556, 557
American Lung Association, 170
American Marketing Association, code of marketing research ethics, 23–25
American Psychological Association, 29
AmEx Company, 9
Analysis of covariance (ANCOVA), 748
Analysis of variance (ANOVA), 735–748
  completely randomized, 741–742
  defined, 737
  dummy-variable regression and, 779–783
  factorial, 744–748
  interaction effect in, 744–748
  latin-square, 743–744
  randomized-block, 742–743
  See also Designs for experiments, statistical
Anchor labels for rating scales, 425–427
Anderson, David R., 489n.4, 492n.5, 701n.5, 702n.6, 716n.10
Anderson, James C., 229n.20, 435n.28
Anderson, Paul F., 815n.1, 818n.6, 840n.15
Anderson, Rolph E., 762n.9, 766n.14
Andreasen, Alan R., 101n.6
Andrews, Frank M., 655n.2, 656n.3
Arbitron Radio Market Reports, 200
Arbitron Television Market Reports, 200
Area sampling, 525–529, 530–531, *table* 534, 557
Artificial-intelligence technologies, 57–59
Assael, Henry, 565n.1
Assmus, Gert, 123n.2
Associates Interviewing Services, 19, 36
Associations, techniques for examining, *table* 688. *See also names of specific techniques*
Attitude
  defined, 414–415
  scaling methods for, 417–420
  self-report measurement of, 420–432
Attribute, 414, 415
Audimeter, 244
Audits and Surveys National Total Market Index, 202
Azhari, Abdul G., 22n.34

Bagozzi, Richard P., 244n.42
Bailar, Barbara, 237n.33
Bailey, Leroy, 237n.33
Balloon test, 226–227
Balsams Grand Resort Hotel, 59
Bankers Life and Casualty Company, 222, *case* 452–460
Banks, Seymour, 311n.1, 321n.9
Bansal, Pradeep, 774n.24
Bar charts, 832–834, *exhibit* 834, *exhibit* 835, 838–839, *exhibit* 840, *exhibit* 842
Barker, R. F., 39n.6
Barnes, Peter, 245n.44
Baty, Wayne Murlin, 818n.6, 819n.7
Baumann, Trish, 50n.12, 52n
Bayesian analysis, 113–124
  decision trees in, 120–122
  defined, 113
  limitations of, 123–124
  market states in, 113–114
  payoff tables in, 114, *table* 115
  posterior probabilities in, 117–118, *table* 119
  prior probabilities in, 114
Behavior, overt, 418
Behavioral variable, 414, 415
BehaviorScan (information service), 55–57, 97, 202
Belden Associates, 579–580
Belief, 414, 415
Belk, Russell W., 817n.1
Bell, Chip R., 38n.4
Bellenger, Danny N., 255n.7, 257n.12, 257n.14, 259, 260n.20, *exhibit* 262n.24, 264–265, 266n.30, 271n.37, 433n.26, 439n.31, 756n.6
Benedict, Thomas C., Jr., *exhibit* 99
Bengston, Roger E., 254n.2
Berenson, Conrad, 236nn.29,30, 577n.5
Bernhardt, Kenneth L., 257n.12, 257n.14, 259, 260n.20, 262n.24, *exhibit* 264–265, 266n.30, 271n.37
Best, Roger, 440n.33
Beta error, 646–647
Bettcher Industries, Inc., *case* 154–156
Between sum of squares, 739
Bishop, George F., 382n.15
Bivariate analysis, 655. *See also* Multivariate techniques
Blair, Ed, 378n.12
Blankenship, A. B., 39n.6, 365n.1
Block, Carl E., 415n.9
Blocking factor, 335, 742
Blue Cross and Blue Shield of Iowa, 98, 99, 132–133
Bluestein, Bram, 12
Blumberg, Leslie, *exhibit* 843
Blunch, Niels J., 372n.5
Bogart, Leo, 30n.50
Bonoma, Thomas V., 133n.3

Booher, Dianna, 816n.2, 818n.6, 819n.7
Boyd, Harper W., Jr., 172n.3, 586
Bradburn, Norman M., 366n.2, 376n.11, 378n.12, 390n.19, 391n.21
Bratton, Gary, 191
Britt, Stewart Henderson, 818
Brownstone, David M., 205
Bruskin, R. H., Associates, 185–187, 188, *table* 565
Buncher, Martin M., 258n.15, 273n.40, 274n.41
Burke Marketing Research, Inc., 18, 255–256, 260n.18, 547
Burton, John, 272n, 372n.6
Bush, Alan J., 240n.38
Bydalek, Nancy, 51

Caballero, Marjorie J., 325n.11
Cadbury, N. D., 290n.11
Calder, Bobby, 256
Callbacks, 579–580
Calvert, Stephen, 327n.13
Camacho, Frank E., 17n, 22n.33
Campbell, Donald T., 216n.4, 246n.46, 287, 295n.20, 302n.22, 311n.1, 325n.10, 328n.14, 331n.18, 442n.35
Campbell Soup Company, 56–57
Cannell, Charles F., 237n.32, 573n.3
Careers in marketing research, 63–69
Carman, James M., 244n.42
Carmone, Frank J., 768n.17
Carruth, Gordon, 205
Cartoon test, *exhibit* 226, 227
Carvey, Davis W., 411n.4
Case, 623
Case study, 132–134
Cattin, Philippe, 775n.27
Causality, conditions for inferring, 283–285. *See also* Experimental research
Cell quotas, determination of, 553
Census Reports, 195–196
Census studies
  conditions for, in marketing, 477–478
  defined, 474
  sampling studies vs., 474–476
Centralization, 16–17
Central limit theorem, 489, 554
Central tendency, 409, 625–626
Channelview Cable Company, 583
Chave, E. J., 433n.24
Cheating by interviewers. *See* Interviewer cheating
Chemical Marketing Research Association (CMRA), 67, 68
Cheseborough Ponds, 290–291
Chicagoland Field Service, *exhibit* 261
Children's Television Workshop (CTW), 210
Child Research Services (CRS), 227, 267, *exhibit* 428
Chi-square contingency test, 689–698
Chi-square goodness-of-fit test, 656–660, 663–664
Chonko, Lawrence B., 25n.38
Churchill, Gilbert A., Jr., 144n.7, 208n.1, 427n.19, 435n.28, 444n.38
Claritas Corporation, 10–11
Classification data, 389–390
Client-participant group, 274
Clients, rights of, 30
Cluster analysis, *table* 736, 765–767
Cluster sampling, 522–525, *table* 534
Coca-Cola Company, 94–95, 432
Code of research ethics, 23–25, 29. *See also* Marketing research ethics
Coding, 614–625
  defined, 101, 614
  errors in, 629
  instructions for, *table* 621
  sheets for, 619, *exhibit* 620, 622
Coefficient of determination (R^2), 708–711
Colgate-Palmolive Corporation, 50
Collins, B. J. K., 330n.15
Common-factor analysis, 762n.9
Communication approach. *See* Questioning methods
Comparative rating scale, 423
Competence, of external research providers, 36
Completely nonstructured question, 216
Completely randomized ANOVA, 741–742
Completely randomized design, 334–335
Completely structured question, 216
Complex question, 386–387
Computer graphics, 841–842
Computerized editing, 612–614
Computerized interviewing, 397–399
  questionnaires for, 397–399
  with telephones, 232–237
Computerized searches for secondary data, 190
Computer languages, fourth-generation, 52–53
Conceptual definition, 415
Conclusive research
  defined, 129
  exploratory research vs., 128–130
  See also Descriptive research; Experimental research
Conditional probabilities of research outcomes, 117
Coney, Kenneth A., 374n.9
Conference Board, 204
Confidence interval, defined, 492
Confidence-interval estimation
  for proportionate stratified random sampling, 516–520
  for simple random sampling, 490–495
Confidence level, 496, 497, 646
Confirmatory research, 129
Confusion matrix, 754, *table* 755
Conjoint analysis, *table* 736, 771–775
  applications of, 775
  defined, 771
  full-profile approach in, 772, 773–774
  graphical interpretation of results of, 774
  trade-off approach in, 772, 773
Conjoint measurement. *See* Conjoint analysis
Consensor, 274
Constant-sum scale, 429
Constitutive definition, 415
Construct validity, 442
Consumer data, sources of, 198–201
Content validity, 442
Contingency coefficient, 694
Contingency table, 689–690, 694–696
Continuing Market Study (CMS) service, 580
Contrived observation, 241–242
Control group, 148, 312. *See also* Experimental research
Controlled research. *See* Experimental research
Controlling (control function), 14
Convenience sampling, 541–544, *exhibit* 545, 546–548
Convergent validity, 442
Cook, Stuart W., 130n.2, 219n.5, 221n.6, 229n.21, 283n.3, 417n.11, 418n.12, 442n.35, 443n.36, 615n.2
Cook, Thomas D., 287n.8, 295n.20, 302n.22
Correlation coefficient
  multiple, 710
  Pearson, 699–704, 710, 729–731
  Spearman, 696–698
Cost, of external research providers, 36

Cost-benefit analysis, 88–92
Covariate, 748
Cover letters for mail questionnaires, 399, *exhibit* 400
Cox, Alan R., 65–66
Cox, Keith K., 272n, 282–283, 305n.25
Cox, William E., Jr., 130, 223n.7
Cramer's V-statistic, 694–695
Crandall, Rick, 26n.40
Cravens, David W., 744n.3
Crawford, C. Merle, 26
Credibility, of external research providers, 35–36
Crespi, Irving, 440n.32
Crisp, Richard D., 15n.13,18
Criterion variable, 706
Cronbach, L. J., 444n.37
Cronbach's alpha, 444n.37
Crosby, Richard, 23
Cross-sectional sample, 137
Cross-sectional studies
  applications of, 138–140
  defined, 137
  longitudinal studies vs., 137
Cross-tabulation, 632–638
Crowded one-on-one interview, 275
Cuadra Associates, 205

Daniells, Lorna M., 204
Darlington, R. B., 720n.12
Data
  primary, 97
  secondary, 97, 131–133
Data accuracy, impact of questionnaire on, 365–367, 388–389. *See also* Errors
Data analysis
  interpretation and, 7–8, 101, 108
  preliminary, 625–638
  role of hypothesis testing in, 654–655
  selecting an appropriate procedure for, 655
  See also Multivariate analysis; Multivariate techniques
Data bank, 46
Data bases, 204–205
  retail, 54–57
  single-source, 57, 58
Data collection, 100–101, 107–108
  forms, 100, 362–404
  system, 46
Data error, 572–575, 584, 587, 588
Data file, 623–625
Data input and retrieval system, 46
Data needs, identification of, 96, 103–104
Data records, 623
Data search, 181, *exhibit* 182
Data set, 623–625, 849–866
Data sources, identification of, 97–98, 104
Datson, Paul G., 225n.12
Davidson, Terrence N., 655n.2, 656n.3
Davies, John O., III, 266n.27, 266n.29
Dawe, Jessamon, 818n.6, 819n.7
Decentralization, 16–17
Decision makers, relationship and communication with researchers, 815–816
Decision making
  based on imperfect information, 118–119
  under certainty, 115, *table* 116
  evolution of systems for supporting, 42–44
  marketing research and, 8–10, 12–15
  under uncertainty, 114–115
Decision rule for rejecting the null hypothesis, 648–650
Decision theory approach, 113. *See also* Bayesian analysis
Decision trees, 120, *exhibit* 121, *exhibit* 122
Dependence techniques in multivariate analysis
  defined, 735
  overview of, *table* 736
  See also names of specific techniques
Dependent variable, in regression analysis, 704, 714–716
Depth interviews, 222–223
Descriptive analysis, vs. inferential analysis, 643
Descriptive research, 209
  conducting, 137–146
  defined, 135
  experimental research vs., 135–137, 282–283
Deshpande, Rohit, 254
Designs for experiments, 310–340
  modifications of standard forms of, 330–338
  pre-experimental, 312–316
  quasi-experimental, 328–330
  statistical, 334–338
  true experimental, 317–328
Deutsch, Morton, 229n.21, 283n.3, 615n.2
Dexter's Department Stores, *case* 592–594
DIALOG Information Service, 204
Dichotomous question, 370
Dichter, Ernest, 254
Diener, Edward, 26n.40
Differential scheduling of interviews, 579–580
Dillman, Don A., 236n.30, 239, 370n.3, 379n.14, 386n.18, 395n.23, 397n.27, 414n.6, 417n.10, 577n.5
Directional hypothesis, 652
Direct observation, 245–246
Directories, as sources of secondary data, 189–190, 203–205
Direct product profitability (DPP), 55
Discriminant analysis, *table* 736, 748–757
  applications of, 751–754, 756–757
  defined, 748
  evaluating a discriminant function in, 754–756
  graphical illustration of, 751
Discriminant axis, 750
Discriminant coefficient, 750
Discriminant function
  defined, 750
  evaluating, 754–756
  using, 751–754
Discriminant score, 750, *exhibit* 751
Discriminant validity, 442
Discriminant weight, 750
Disguised observation, 242–243
Disguised question, 219–220
Dispersion, measures of, 625–626
Disproportionate stratified random sampling, 520–522
Dolan, Dennis J., 45n.5
Don't know response, 373–374, 616, 618
Double-barreled question, 379–380
Dual-moderator group, 273
DuBoff, Robert S., 610n.1
Dueling-moderator group, 273
Dummy variables
  defined, 779
  regression analysis with, 779–783
Dun & Bradstreet, Inc., 204
Dunkelberg, William C., *exhibit* 611, 618n.4
du Pont, F. George, 266n.28
DuPont, Thomas D., 290n.14
Durand, Richard M., 828n.11
Dutka, Solomon, 683n.1

Editing, 101, 608–614
Eigenvalues, in factor analysis, 762
Eisenhart, Tom, 48n, 57n.22
Elbert, Norbert F., 441n.34
Electronic observation, 212
Eliashberg, Jehoshua, 764n.11
Elrick and Lavidge, Inc., 262n.22, 631
End piling, 425
Enis, Ben M., 282–283, 305n.25
Enrico, Roger, 94
Erdos, Paul L., 395n.23, 399n.29, *exhibit* 400

Error mean square, 713
Errors
  beta, 646–647
  data, 572–575, 584, 587, 588
  nonresponse, 570–572, 577–583, 588
  nonsampling, 563–587, 588
  sampling, 485–486, 512, 563–567, 588
  sampling frame, 567–570, 575–577, 588
  systematic, 567
  in written reports, 828–829
Error sum of squares, 708
Ethics. *See* Marketing research ethics
Expected value
  under certainty, 115–116
  contingent, *table* 120
  of decision making with imperfect information, 118–119
  of imperfect information, 119–120
  of perfect information, 116–117
  under uncertainty, 114–115
Experiment
  choice of, 289–295
  defined, 282
  potential biases in, 295–306
  types of, 285–288
Experimental designs. *See* Designs for experiments
Experimental groups, 148, 312
Experimental research, 209, 280–307
  causation in, 283–285
  conducting, 146–148
  defined, 135
  descriptive research vs., 135–137, 282–283
  external validity in, 287, 302–306
  internal validity in, 287, 295–302, 305–306
  laboratory vs. field experiments in, 285–288
Expert-opinion survey, 130–131
Expert systems (ES), 43, 57–59
Explanatory variable, 704
Exploratory research, 209
  conclusive research vs., 128–130
  conducting, 130–135
  defined, 128
External providers of marketing research, 15, 17–19, *exhibit* 36, *exhibit* 37
  guidelines for using, 35–39
  See also names of specific firms
External secondary data, 180, 181–190, 195–205
External validity
  defined, 287
  threats to, 302–306
  See also Validity
Eye-tracking equipment, 243

Face validity, 442
Factor analysis, *table* 736, 757–764
  achieved communality in, 762
  applications of, 763–764
  defined, 748
  eigenvalues in, 762
  factor loading in, 759
  factor-loading matrix in, 760–761
  factor rotation in, 763n.10
  factor scores in, 759
  graphical illustration of, 758, 760
  interpretation of results of, 760–763
Factorial ANOVA, 760–763
Factorial design, 337–338
Falthzik, Alfred M., 382n.15
Fantastic Car Wash, *case* 806–811
Farris, Paul W., 290n.13
Fast-Test service, *case* 158–162
Ferber, Robert, 29n.47, 145nn.8,10, 429n.21, 625n.6, 753n.4, 764n.12
Fern, Edward F., 257n.13
Ferrell, O. C., 23n.36, 815n.1
Field edit, 608–609
Field experiment
  defined, 286
  laboratory experiment vs., 285–288
Field interviewers, ethics for, 25
Field service firms, rights of, 28–29
Fierman, Jaclyn, 95n
Filter question, 385–386, 394
Final edit, 609–610
FINDSEND service, 190
FIND/SVP (marketing research firm), 190
Finn, Nancy, 28
Fiske, Donald W., 442n.35
Focus group, 131, 255–273
  advantages of, 266–268
  applications of, 270–273
  client-participation, 274
  composition of, 257–260
  conducting, 263–266
  defined, 255–256
  disadvantages of, 269–270
  dual-moderator, 273–274
  dueling-moderator, 273–274
  moderator of, 260–266
  respondent-moderator, 274
Foot-in-the-door technique, 581n.9
Forced-choice scale, 423–424
Ford Motor Company, 6–7, 14, 87
Foreign statistics, government sources of, 197–198
Fourth-generation languages (4GLs), 52–53
Frank, Ronald A., 766n.14, 767n.15
Frazer, Scott C., 581n.9
Freedman, Jonathan L., 581n.9
Friedman, Hershey H., 422n.15
Friedman, Linda W., 422n.15
Fries, James R., 191n.17
Frito-Lay, Inc., 50
F-test, 710–711, 719
Fuerst, William L., 52n.15
Full-profile approach in conjoint analysis, 772, 773–774
Funnel sequence of questions, 390–392

Gabor, Andre, 292n.17
Gallagher, William J., 817, 818n.6, 827–828
Gallup Organization, Inc., 201, 388, 556, 558
Galvanic skin response (GSR), 419
Garreau, Joel, 174n.5
Gates, Roger, 240n.36
General public, rights of, 31
Gerbing, David W., 435n.28
Geurts, Michael D., 378n.12
Gibbons, Jean Dickinson, 655n.1, 662n.5, 695n.3
Glasser, Gerald, 235n.26, *exhibit* 491
Goldman, Alfred E., 257n.12, 266n.30
Goldstucker, Jac L., 255n.7, 256n.12, 257n.14, 259, 260n.20, *exhibit* 264–265, 266n.30, 271n.37, 433n.26, 439n.31, 756n.6
Goodness-of-fit test, chi-square, 656–660, 663–664
Government sources of secondary data, 183, 195–198, 203–204
Granger, Clive W., 292n.17
Graphic rating scale, 421–422
Graphics, for reports, 829–840
Green, Paul E., 124n.4, 429n.21, 766n.14, 767n.15, 768nn.17,18, 771nn.21,22, 772n.23, 774n.24, 775n.25
Greenbaum, Thomas L., 257n.12
Greenberg, Barnett A., 255n.7, 433n.26, 439n.31, 756n.6
Greenberg, M. G., 429n.21
Greer, Thomas V., 130n.1
Gubernick, Lisa, 272n.38
Guenther, William C., 741n.1, 742n.2
Guffey, Hugh J., Jr., 828n.11
Gundee, Howard N., 240n.37
Gunning, Robert, 827n.9

Hair, Joseph F., Jr., 240n.38, 762n.9, 766n.14
Haire, Mason, 227n.19
Haley, Douglas F., 59
Haley, Russell I., 771n.20
Hamburg, Morris, 656n.3, 682n.2
Hamilton Bank, *case* 349–350, *case* 798–800
Hanes Company, 13–14, 546–547
Harman, Harry H., 757n.8, 763n.10
Harris, Catherine L., 213n.2
Harris Survey, 201, 473, 586
Harry Heller Research Corporation, 30
Hartley, Robert F., 7n.1
Hassett, Matthew, 669n.6
Hatoun, Antoine, 54n.17
Hawkins, Del I., 374n.9, 440n.33
Haynes, J. B., 38n.5, 39n.7
Heller, Harry, 30
Hershey Chocolate Company, *case* 353–356, *case* 461–469
Hess, John, 266n.30
Higginbotham, James B., 272n
Himstreet, William C., 818n.6, 819n.7
History effect, 295–296
Hit rate, 754–755
Hogue, Jack T., 50n.13
Holiday Inns, 46–47
Holland, Charles W., 744n.3
Hollander, Sidney, Jr., 29n.47
Holloway, Robert J., 318n.5
Home Testing Institute (HTI), 140, 142
Honomichl, J. J., 11n.5, 64n.2
Hooke, Robert, 683n.1
Hospitality Standards, 275, 276
Houston, Michael J., 130n.1, 399n.28
Huff, Darrell, 631n.8, 633n.10, 835n.12
Hughes, G. David, 424n.16, 433n.25
Human observation, 243
Hunt, Shelby D., 25n.38, 396n.25, 397n.26
Hybrid-sampling plan, 556
Hypothesis testing, 643–678
  with chi-square goodness-of-fit test, 656–660, 663–664
  errors in, 645–647
  with Kolmogorov-Smirnov test, *table* 656, 660–662, 663–664
  one-tailed, 652–653
  overview of, 643–654
  role in data analysis, 654–655
  for single mean, *table* 656, 662–667
  for single proportion, *table* 657, 665–667
  steps in, 654
  of two means (dependent samples), *table* 657, 672–674
  of two means (independent samples), *table* 657, 667–672
  of two proportions, *table* 657, 674–677
  two-tailed, 653–654

Ideal population, 542–543, *exhibit* 545, *exhibit* 546
Implied population, 542–543, *exhibit* 545, *exhibit* 546
Incidence rate, 501
Independence techniques in multivariate analysis
  defined, 735
  overview of, *table* 736
  See also names of specific techniques
Independent variable in regression analysis, 704, 714–716
In-depth interviews, 222–223
Indexes, as sources of secondary data, 189–190, 203–205
Indirect observation, 246
INF*ACT (data base), 201
Inferential analysis, vs. descriptive analysis, 643
Information, defined, 8
Information consultants, 191
Information dissemination system, 46
Information Resources, Inc., 55, 97
Infosource, Inc., 191
In-house marketing research, 15–17
Institute for Scientific Information, 205
Instrument variation effect, 298–299
Interaction effect, 338, 744–748
Internal secondary data, 180–181
Internal validity
  defined, 287
  threats to, 295–302, 305–306
  See also Validity
Interval, 408
Interval scale, 410–412
Interviewer cheating, 573
  detection of, 585
  minimization of, 586
Interviewer ethics, 25
Interviews
  computer-assisted telephone, 232–237
  computerized, 232–237, 397–399
  depth, 222–223
  focus group. *See* Focus group
  openers for, 399–401
  personal, 229, 230, 399–401
  shopping mall intercept, 230, 240, *exhibit* 241
  telephone, 229–240, 399–401
Inverted-funnel sequence of questions, 391–392
Itemized rating scales, 421–422
ITT, *case* 789–791

Jahoda, Marie, 229n.21, 283n.3, 615n.2
Johansson, Johny K., 254n.4
Johnson, D. K., 254n.4
Johnson Company, 272
Johnston, J., 705n.7
Jolson, Marvin A., 382n.15
Joselyn, Robert W., 93n.2, *exhibit* 262n.23, 264–265
Judgment sampling, 544–548

Kahan, Hazel, 260n.19
Kahn, Robert L., 573n.3
Kamen, Joseph M., 22n.34
Kanuk, Leslie, 236nn.29,30, 577n.5
Katz, Gerald M., 292n.17, 305n.24
Keon, John, 565n.1
Kerin, R. A., 35n.1
Kerlinger, Fred N., 283n.2, 286n.7, 302–303, 318, 721n.13, 723n.14, 782n.1
Kessler, Felix, 57n.20
Key-informant technique, 130–131
Kidder, Louise H., 317, 318n.7, 330n.17
King, R. L., 266n.30
Klein, Laura, 655n.2, 656n.3
Klocke, William J., 176n.7
Knain, D. Matthew, 17n, 22n.33
Kolmogorov-Smirnov test, *table* 656, 660–662, 663–665
Korgaonkar, Pradeep K., 774n.24
Kotler, Philip, 44n.4, 54n.16, 123n.3, 290n.11
Krieger, Abba M., 774n.24
Kruskal, J. B., 769n.19
Kruzas, Anthony T., 205
Kurnow, Ernest, *exhibit* 491

Labeled response choices, 425–427
Laboratory experiment
  defined, 286
  field experiment vs., 285–288
Labovitz, Sanford, 411n.5, 431n.22
Laczniak, Gene R., 815n.1
Lamont, Lawrence M., 435n.28
Landis (Hauser), Marilyn, 26n.39
Lang, Peggy, 263n.25
Languages, fourth-generation, 52–53

Latin-square analysis of variance, 743–744
Latin-square design, 336–337
Laurent, Andre, 237n.32
Lavidge, Robert L., 631
Lawson, Leigh, 578n.7
Leading question, 380–381
Least-squares approach in regression analysis, 705–707, 731–733
Legal Marketing Research, Inc., 18–19
Lerner, David, 139n.4, 330n.15
Levi Strauss & Company, 13
Libresco, Joshua D., 20n.24
Lieberman, David, 245n.44
Likert, Rensis, 433
Likert scale, 433–437
Limited, The, 212
Line charts, 831, *exhibit* 832, *exhibit* 842
Liniger, Charles A., 373n.8, 382n.15
Linsky, Arnold S., 236nn.29,30, 577n.5
Lipman, Joanne, 245n.44
Litmus Simulated Test Marketing (LSTM) service, 241–242, 292–294
Little, John D. C., 54n.17
Loaded question, 380–381
Locander, William B., 372n.6
Lodish, Leonard, 291
Loether, Herman J., 630n.7, 633n.10, 717–718
Longitudinal studies
  applications of, 140–146
  cross-sectional studies vs., 137
  defined, 137
Lord, William Jackson, Jr., 818n.6, 819n.7
Ludwig, Jacob, 379n.14, 391n.20
Lundstrom, William J., 435n.28
Lusch, Robert F., 815n.1

McCaw Cablevision, *case* 601–602
McCollum/Spielman Associates, Inc., 227, 267, *exhibit* 428
McDaniel, Stephen W., 476n.3
McDonald, Susan Schwartz, 257n.12
McElnea, Jeffrey K., 330n.16
McGronagle, John J., Jr., 38n.4
MacLachlan, James, 244n.42
McLeod, Raymond, Jr., 44n.3, 49nn.9,10
McNemar test, 694n.1
McTavish, Donald G., 630n.7, 633n.10, 717–718
Madden, Charles S., 476n.3
Mail surveys
  designing cover letters for, 399, *exhibit* 400
  personal interviews and telephone surveys vs., 229–240
  reducing nonresponse errors in, 577–578
Main, Jeremy, 212n
Main effect, 741, 746–748
Manipulation checks, 312
Market Facts, Inc., 232, *exhibit* 233, 255n.8
Marketing concept, 13
Marketing decision support system (MDSS), 43, 49–57
  with access to retail data bases, 54–57
  applications of, 51–52
  defined, 49
  distinguishing features of, 50
  fourth-generation languages and, 52–53
  marketing response functions in, 53–54
Marketing information system (MkIS), 43, 44–49
  components of, *table* 46
  defined, 44
  developing and using, 48
  structure of, 45
Marketing manager, interface with researchers, 19–22
Marketing mix, 13
Marketing plans
  developing and implementing, 13–14
  evaluating, 14–15
Marketing research
  ad hoc, 41, 43
  applications of, 5, 10–12
  careers in, 63–69
  defined, 5, 15
  external providers of, 15, 17–19, 35–39
  in-house, 15–17
  key characteristics of, 6–12
  organization of, 16–17
  role in decision making, *exhibit* 10, 12–15
  systems-oriented, 42
Marketing researchers, interface with marketing manager, 19–22
Marketing research ethics, 23–31
  with clients, 30
  code of, 23–24
  defined, 23
  for field interviewers, 25
  with field service firms, 28–29
  with general public, 31
  for research practitioners, 24–25
  with research subjects, 29–30
  with research suppliers, 27–28
  for users of marketing research, 25
Marketing research firms. *See* External providers of marketing research
Marketing research jobs, 63–69
Marketing research project
  analyzing data and interpreting results in, 101, 108
  choosing an appropriate design for, 98–100
  data collection in, 100–101, 107–108
  defining purpose of, 92–96, 103
  developing data collection forms for, 100
  evaluating, 85–92, 102–103, 109
  identifying data needs in, 96, 103–104
  identifying data sources for, 97–98, 104
  illustration of, 102–109
  sample design in, 100, 105–107
  stages of, 86
  steps in, 85–110
Marketing research users, ethics for, 25
Marketing response functions, 53–54
Market states in Bayesian analysis, 113–114
Marquis, Kent H., 237n.32
Marriott Corporation, 16–17
Martilla, John A., 411n.4
Martin, Claude R., Jr., 244n.43
Martin, Warren S., 427n.19
Mastic Corporation, 48
Matching, 317–319
Matheson, Richard E., 275n.43
Maturation effect, 296–297
Maximum chance criterion, 756
Mayer, Charles S., 37n.3, 395n.22, 625n.6
Mayros, Van, 45n.5
Mazis, Michael B., 426n.18
Mean, 411
Mean square, 711
Measurement, 406–446
  of attitudes, 417–432
  defined, 407
  multiple-item scales, 433–445
Measurement error. *See* Data error
Measurement levels of data, 408–413, 429–431, 619, *table* 626
Mechanical observation, 243
Median, 409–410
Menezes, Dennis, 441n.34
Metric data, 413
Metzger, Gale D., 235n.26
Michand, Bridgid J., 775n.28
Milliman, Ronald E., 169
Mindak, William A., 437n.30

Mirro Company, 14–15
Missing-value category, 618
Mizerski, Richard W., 327n.13
Moberg, Patricia E., 235nn.26,27, 568n.2, 581n.10
Mode, 409
Models/techniques system, 46
Modern Silo Systems, *case* 75–77
Mohan, Lakshmi, 54n.17
Montgomery Ward, 12
Moore, William L., 775n.25
Morrison, Donald F., 753n.4, 754n.5
Mortality effect, 300–302
Morton, John, 757n.7
Motivation research, 223–229
MRI Mediamarketing Service, 200–201
Mullet, Gary M., 411n.4
Multicollinearity in regression analysis, 720–723
Multidimensional scaling, *table* 736, 767–771
Multiple-category question, 370–371
Multiple correlation coefficient, 712
Multiple-item scale, 433–445
  single-item scale vs., 433, 441–445
  strengths of, 441–445
  types of, 433–441
Multiple regression analysis, *table* 688, 704, 718–723. *See also* Regression analysis
Multistage area sampling, 527–528, 530–531
Multivariate analysis, 655
Multivariate techniques
  dependence, 704–723, 735–757, 779–783
  independence, 735, 757–775
  overview of, *table* 736
  See also names of specific techniques
Murphy, Patrick E., 815n.1
Murray, Henry A., 225n.13
Myers, James H., 37n.3, 90n.1, 411n.4, 426n.17
Myers, John G., 244n.42

Nadler, Leonard, 38n.4
Namelab (company), 18
Narayana, Chem L., 422n.15
National Center for Health Statistics, 389
National Family Opinion, Inc. (NFO), 139–140, *exhibit* 141
National Insurance Company, *case* 799–805
National Purchase Diary Panel, Inc. (NPD), *exhibit* 142, 145–146, 172, 198–200, 552, *table* 553, 577, *data set* 849–866
Natural observation, 241–242
NBC, 31
Nelson, Ronald, 244n.43
Nestle Foods Corporation, 59
Neuringer, Charles, 225n.14
Nevin, John R., 288n.10, 399n.28
Newman, Joseph W., 113n.1, 227n.17
Nielsen, A. C., Company, 18, 19, 36, 57, 97, 172, 183, 200, 201, 244, *exhibit* 245, 294–295, 332–333, 473, 521, *exhibit* 522, 612–614
Nielsen New Store Quality Edit, 612
Nielsen Retail Index, 183, 201, 473, 474, 521, *exhibit* 522
Nielsen Scantrack Service, 57, 97, 172, 201, 295, 612–614
Nielsen Store Usability Edit, 612–614
Nielsen Television Index (NTI), 200
Nighswonger, Nancy J., 244n.43
Nissan Motor Company, 9
Nominal scale, 408–409
Nonaka, Tkujiro, 254n.4
Noncomparative rating scale, 423
Nondirectional hypothesis, 652–653
Nondisguised observation, 242–243
Nonforced-choice scale, 423–424
Nonlinear association, 703
Nonmetric data, 413
Nonparametric procedures
  chi-square goodness-of-fit test, 656–660, 663–664
  defined, 655
  Kolmogorov-Smirnov test, *table* 656, 660–662, 663–664
  types of, *table* 656
Nonprobability sampling
  defined, 478
  probability sampling vs., 478–480, 553–556
  types of, 540–559
Nonrepresentative-sample bias, 304–305
Nonresponse error, 570–572, 577–583, 588
  control of, 577–582, 588
  defined, 570
  incorporating effects into sample results, 582–583
  sources of, 571–572
Nonsampling errors, 563–564, 588
  controlling, 575–587, 588
  defined, 563
  sampling errors vs., 564–567
  types of, 567–575
Nonstructured, disguised questionnaires, 223–229
Nonstructured, nondisguised questionnaires, 221–223
Nonstructured observation
  qualitative research and, 251
  structured observation vs., 246–247
Nonstructured questions, 367, 369–370
  qualitative research and, 252
  structured questions vs., *table* 218
Normal probability distribution, 489–490, 554
Northwest Bank of Woodcreek, *case* 595–597, *case* 786–788
Not-at-home problem, 571, 578–581
NPD Research, Inc., *exhibit* 142, 145–146, 172, 198–200, 552, *table* 553, 577, *data set* 849–866
Null hypothesis
  decision rule for rejecting, 648–650
  statement of, 645
Numerical codes, assigning, 618–623
Nunnally, Jum C., 428n.20, 442n.35, 757n.8, 763n.10

Observation
  defined, 208, 623
  designing forms for, 401–403
  electronic, 212
  questioning vs., 209–216
  types of, 240–247
Office edit, 609–610
O'Hara, James F., 372n.7
O'Malley, Patrick M., 655n.2, 656n.3
Omnibus panel, 139, *exhibit* 141, *exhibit* 142
One-group, after-only design, 313–314
One-group, before-after design, 314–315
One-sided question, 381
One-stage area sampling, 525–526
One-stage cluster sampling, 523–525, *table* 534
One-tailed hypothesis test, 652–653
One-way tabulations, 627–631, 835–839
Open-ended questions, 217, 367, 369–370. *See also* Nonstructured questions
Openers for personal and telephone interviews, 399–401
Operational definition, 415–416
Opinion Research Corporation, 11, 478, 479–480, 530–531

- Oppenheim, A. N., 373n.8
- Oral presentations, 841–843
- Ordinal scale, 409–410
- Original source, 179–180
- Osgood, Charles E., 437
- Ottman, Frederick R., *exhibit* 491
- Overt behavior, observing, 418

- **Pacific Northwest Bell Telephone Company,** 568, 581
- Paired-comparison rating scales, 429, 431, 432
- Panel
  - defined, 137
  - omnibus, 139, *exhibit* 141, *exhibit* 142
  - true, 144–146
- Panel conditioning, 145
- Panel design, 328–330
- Parameter, 485
- Parametric procedures
  - defined, 655
  - test for a single mean, *table* 656, 662–665
  - test for a single proportion, *table* 657, 665–666
  - test of two means (dependent samples), *table* 657, 672–674
  - test of two means (independent samples), *table* 657, 667–672
  - test of two proportions, *table* 657, 674–677
  - types of, *table* 656–657
- Parasuraman, A., 21n.29, 64nn.1, 3, 66n.6, 68nn.8,10, 69n.12, 255n.9, 578n.7, 756n.6, 815n.1
- Parfitt, J. H., 330n.15
- Partial correlation analysis, 721n.13
- Paul, R., 35n.1
- Payne, Melanie S., 267n.31
- Payne, Stanley L., 366n.2, 371n.4, 378n.13, 383n.17, 396n.24
- Payoff table in Bayesian analysis, 114, *table* 115
- Pearson correlation coefficient, *table* 688, 699–704, 710, 729–731
- Pedhazur, Elazar J., 723n.13, 725n.14, 784n.1
- Pekelman, Dov, 291
- People meter, 244, *exhibit* 245
- People's National Bank, *case* 450–451
- **Perception Research Services, Inc.,** 420
- Perdue, Barbara C., 312n.2
- Performance evaluation, 419
- Periodicity in systematic sampling, 532
- Perreault, William D., Jr., 411n.5
- Personal interviews, 229
  - openers for, 399–401
  - reducing nonresponse errors in, 578–582
  - telephone and mail surveys vs., 229–240
- Pessemier, Edgar A., 418n.13
- Peter, J. Paul, 427n.19, 442n.35, 444n.37
- Petersen, R. A., 35n.1
- Peterson, Greg, 51–52
- **Phoenix Newspapers, Inc.,** *case* 598–600
- Physiological responses, monitoring, 419–420
- Pictograms, 836–837, *exhibit* 838
- Pictorial profile, 438, *exhibit* 439
- Picture labels for rating scales, *exhibit* 428
- Pie charts, 830–831, *exhibit* 842
- Piper, Cindy, 395n.22
- Planchon, John M., 828n.11
- Plus-one dialing, 576
- Point estimate from regression analysis, 716
- Poltrack, David F., 245n.44
- Pond, Dale, 21
- Pooled standard deviation, 670
- Pope, J., 273n.39
- Population
  - defined, 474
  - ideal, 542–543, *exhibit* 545, *exhibit* 546
  - implied, 542–543, *exhibit* 545, *exhibit* 546
- Population parameter value, relationship of power to, 683–686
- Posterior probabilities in Bayesian analysis, 117–118, *table* 119
- Power
  - defined, 646–647
  - significance level and, 683–686
- Precision level, 496, 497
- **Predicasts, Inc.,** 203, 205
- Prediction error, 731, *exhibit* 732
- Prediction interval from regression analysis, 716
- Predictive validity, 443
- Predictor variable, 704
- Pre-experimental designs, 312–316
  - features of, 312–313
  - one-group, after-only form of, 313–314
  - one-group, before-after form of, 314–315
  - two-group, *ex post facto* form of, 315–316
- Preliminary data analysis, 625–638
  - measures of central tendency and dispersion in, 625–626
  - one-way tabulation in, 627–631
  - two-way tabulation in, 631–638
- Preliminary edit, 608–609
- Presser, Stanley, 379n.14, 391n.20
- Pressley, Milton M., 236n.31
- Prestige, of external research providers, 37–38
- Pretesting effect, 297–298, 303
- Pretesting questionnaires, 396–397
- Pretest-manipulation interaction bias, 303–304
- Pride, William M., 325n.11
- Primary data, 97, 207–248
  - collection methods for, 207–209
  - defined, 170, 207
  - limitations and advantages of, relative to secondary data, 170–180
  - observation techniques for, 240–247
  - questioning vs. observation, 209–216
  - questionnaire administration methods for, 229–240
  - questionnaire format for, 216–229
- Principal-components analysis, 764n.9
- Pring, David C., *exhibit* 239
- Prior probabilities in Bayesian analysis, 114
- Private guides to secondary data, 204–205
- **PRIZM** (information base), 11, 202–203
- Probability-proportional-to-size area sampling, 528–529
- Probability sampling
  - comparison of methods of, 533–535
  - defined, 478
  - nature of, 479–480
  - nonprobability sampling vs., 478–480, 553–556
  - types of, 508–536
- Problem definition, 92–96, 103
- **Procter and Gamble,** 290–291
- Product-moment correlation. *See* Pearson correlation coefficient
- Professional respondents, 260
- Professional Ski Supply Company, *case* 357–358
- Projective techniques, 224–229
- Proportional chance criterion, 758
- Proportionate stratified random sampling, 510–520
  - confidence-interval estimation in, 516–520
  - sampling distribution for, 512–516
  - sampling error in, 512
  - simple random sampling vs., 512–514
  - steps in, 510

Pseudo-research, 22, 87
Public-use micro-data files, 183, *exhibit* 184
Punj, Girish, 767n.16
Purposive sampling, 544–548

Q-factor analysis, 766n.14
Quaker Oats Company, 50–52
Qualitative principles, 6–7
Qualitative research, 250–278
  applications of, *table* 253
  defined, 251
  importance of, 252–255
  quantitative research vs., 252
  techniques for, 273–275
  See also Focus group
Quantitative research
  applications of, *table* 252
  qualitative research vs., 252–253
Quasi-experimental designs, 328–330
QUBE (data collection system), 231n.22
Question
  complex, 386–387
  dichotomous, 370
  disguised, 219–220
  double-barreled, 379–380
  filter, 385–386, 394
  form of, 367–376
  with implicit assumptions, 383–386
  leading, 380–381
  multiple-category, 370–371
  one-sided, 381
  open-ended, 217, 367, 369–370
  relevance and wording of, 376–387
  sequencing of, 387–394
  structured, 216–219, 367, 370–376
Questioning methods
  defined, 208
  observation vs., 209–216
Questionnaire
  appearance and layout of, 394–396
  for computerized interviewing, 397–399
  cover letter for, 399, *exhibit* 400
  data accuracy and, 365–367, 388–389, 584, 587, 588
  defined, 208, 363
  as link between researcher and respondent, *exhibit* 366
  pretesting of, 396–397
  types of, 220–229
Questionnaire administration methods, 229–240
  accuracy of, 234–238
  choosing, 238–240
  comparison of, *table* 238
  control over, 234–238
  cost of, 232–234
  popularity of, *table* 230
  questionnaire design and, 375–376
Questionnaire design, 104–105, 363–399
  complexity of, 364–365
  elements of, 367, *exhibit* 368
Questionnaire format, 216–229
  disguise and, 219–220
  structure and, 216–219
  types of, 220–229
Quota sampling, 548–553, 557

Rabin, A. I., 225n.12
Ramond, Charles, 284, 319n.8
Rand Corporation, *table* 483
Random assignment, 317–320
Random-digit dialing, 576
Randomized-block ANOVA, 742–743
Randomized-block design, 335–336
Random numbers, *table* 483
Random sampling, 318
  disproportionate stratified, 520–522
  proportionate stratified, 510–520
  simple, 480–495, 502, 512–514
  stratified, 509–510
Random selection, 479
Rank order, 408, 432
Rao, Vithala R., 768n.17
Ratio, 408
Ratio scale, 412–413
Raw variables, 624
Ray, Michael L., 214n.3, 242
Reactive bias, 302–303
Reagan, Ronald, 11
Recoded variable, 624–625
Regression analysis, 704–723, *table* 738
  applications of, 713–714
  dependent variables in, 704
  with dummy variables, 779–783
  graphical illustration of, 732
  independent variables in, 704
  least-square approach for, 705–707, 731–733
  multicollinearity in, 720–723
  multiple, *table* 688, 704, 718–723
  point estimate from, 716
  precautions in using, 714–718, 720–723
  prediction interval from, 716
  R^2, 708–711
  simple, *table* 688, 704–718
  slope in, 707, 711–713
  standard error in, 711
  standardized coefficient in, 713
  Y-intercept in, 707
Regression equations, 705–714
  deriving, 705–707
  evaluating, 707–713
  practical applictions of, 713–714
Regression mean square, 711
Regression sum of squares, 708
Reibstein, David J., 290n.13
Reinmuth, James E., 378n.12
Reliability, 443–444
Reports
  oral, 841–843
  written, 816–840
Research continuum, 135
Research design, choice of, 98–100
Research practitioners, ethics for, 24–25
Research process, 85, *exhibit* 86. *See also* Marketing research project
Research proposal, 98–100
Research results, presentation of, 815–844
Research subjects, rights of, 29–30
Research suppliers, rights of, 27–28
Residual mean square, 711
Residual sum of squares, 708
Respondent-moderator group, 274
Respondent refusal problem, 571–572, 581–582
Response categories, 371–376, 615–618
Response category sequence rotation, 371–372
Response distributions, skewed, 630
Response functions, 53–54
Response latency, 243–244
Retail data, sources of, 201–202
Retail data bases, 54–57
Reynolds, F. D., 254n.4
Reynolds, R. J., Tobacco Company, 65–67
R-factor analysis, 766n.14
Riche, Martha Farnsworth, 203n.2
Robbins, Jonathan, 11
Robertson, Thomas S., 764n.11
Robinson, Patrick J., 767n.15, 771n.20
Rodgers, Willard L., 655n.2, 656n.3
Roering, Kenneth J., 415n.9
Rogers, John C., 44n.3, 49n.9
Roper Organization, Inc., 185, 201, 557
Rothe, J. T., 39n.7
Rubel, Thomas I., 174nn.4,5
Rubinov, Steve, 51
Ruby Restaurant, *case* 792–795
Rudelius, William, 582n.11
Runyon, Kenneth, 139n.5
Russick, Bertram, 582n.11
Rypka, Karen, 11n.4

St. John Bate, Joseph, 46n.6, 49n.8, 52n.14
Salaries, 64, *exhibit* 65
SAMI-Burke, 57, 202
SAMI reports, 202
Samli, A. Coskun, 90n.1
Sample design, 100, 105–107. *See also* Nonprobability sampling; Probability sampling
Sample size, determination of, 495–502
Sampling, 472–504
  area, 525–529, 530–531, *table* 534, 557
  census study vs., 474–476
  cluster, 522–525, *table* 534
  convenience, 541–544, *exhibit* 545, 546–548
  defined, 473–474
  hybrid, 556
  judgment, 544–548
  probability vs. nonprobability, 478–480
  quota, 548–553, 557
  random. *See* Random sampling
  systematic, 529–533, *table* 535
  See also Nonprobability sampling; Probability sampling
Sampling distribution, 486–490, *exhibit* 491
  defined, 487
  for proportionate stratified random sampling, 512–516
  for simple random sampling, 488–489
Sampling error, 485–486, 564, *table* 565
  defined, 485, 563
  estimation of maximum, 566
  minimizing, 588
  nonsampling errors vs., 564–567
  in proportionate stratified random sampling, 512
Sampling frame, 474
Sampling frame error, 567–570
Sampling interval in systematic sampling, 529
Sarin, Rakesh Kumar, 54n.16, 123n.3
Sawyer, Alan G., 292n.18
Scales
  of measurement levels of data, 408–413
  for measuring attitudes, 420–432
ScanAmerica service, 202
ScanTrack (information service), 57, 97, 172, 201, 295, 612–614
Scatter diagram
  defined, 700
  illustrations of, 700, 701, 703, 708
Scheibenreis, Don, 52
Schlaifer, Robert, 113n.1
Schmitt, Paul M., 57, *exhibit* 58
Schuman, Howard, 379n.14, 391n.20
Schwartz, David, 244n.43
Schwartz, Richard D., 216n.4, 246n.46
Sears, 22, 191–192
Sechrest, Lee, 216n.4, 246n.46
Secondary data, 97, 168–205
  abstracts, directories, and indexes as sources of, 189–190, 203–205
  accuracy of, 176–177
  advantages of, 170–172
  analysis of, 131–133
  computerized searches for, 190
  defined, 169
  external, 180, 181–190, 195–205
  government sources of, 183, 195–198, 203–204
  internal, 180–181
  limitations of, 172–180
  miscellaneous sources of, 189
  relevance of, 172–176
  sources and types of, 180–190
  syndicated sources of, 183–187, 198–203
  trade associations as sources of, 187–189
Secondary-data management, 191–192
Secondhand source, 179–180
Selection effect, 300
Self-report measurements of attitudes, 420–432
Selling Areas-Marketing, Inc. (SAMI), 57, 202
Selltiz, Claire, 130n.2, 219n.5, 221n.6, 229n.21, 283n.3, 417n.11, 418n.12, 442n.35, 443n.36, 615n.2
Semantic-differential scale, 437–439
Semeonoff, Boris, 224n.10
Sendak, Paul E., 292n.18
Sensitivity, 444–445
Sentence completion test, 224–225
Sequencing, of questions, 387–394
"Sesame Street," 210
Sheth, Jagdish N., 764n.12, 776n.29
Shiffler, Ronald E., 499n.6
Shimp, Terence A., 815n.1
Shop 'n Save, 55, 59
Shopper study, 275–276
Shopping mall intercept interviewing, 230, 240, *exhibit* 241
SIC codes, 196n.1
Sidel, Philip S., 625n.6
Sid's Sandwich Shops, *case* 351–352
Siegel, Sidney, 655n.1, 662n.4, 694nn.1,2, 698n.4
Sigelman, Lee, 391n.20
Significance level, 646–652
  choice of, 647–648
  confidence level and, 646
  defined, 646
  power and, 683–686
  in two-tailed hypothesis test, 653
Simmons Market Research Bureau, Inc., 201
Simmons Media/Marketing Service, 201
Simon, Julian L., 132, 133n.3, 145n.8
Simple cluster sampling, 523–525, *table* 534
Simple random sampling
  choosing a sample with, 482–485
  confidence-interval estimation in, 490–495
  defined, 480–482
  limitations of, 502
  proportionate stratified random sampling vs., 512–514
  sample distribution for, 488–489
Simple regression analysis, *table* 688, 704–718
Simulated test markets, 292–294
Single-item scale, 433, 441–445
Single-source data base, 57, 58, 201
Skewed response distributions, 630
Skinner, Steven J., 23n.36
Skip patterns in questionnaires, 392–394
Slides, *exhibit* 842–843
Slope in regression analysis, 707, 711–713
Smith, George Horsley, 223n.9, 224n.11, 225n.15, 226n.16
Smith, Joan Macfarlane, 223
Smith, S. A., 22n.31
Solomon, Paul J., 240n.36
Solomon, R. L., 326n.12
Sonquist, John A., *exhibit* 611, 618n.4
Sowter, Anthony P., 292n.17
Sparkman, Richard D., Jr., 396n.25, 397n.26
Spear, Mary E., 840n.15
Spearman correlation coefficient, *table* 689, 696–698
Split-ballot technique, 372, 382–383
Split-half reliability, 443–444
SPSS, hypothesis tests using, 663–664
SPSS computer output
  for ANOVA analysis, *exhibit* 740
  for dummy-variable regression, *exhibit* 781

- for multiple regression analysis, *exhibit* 719
- for pairwise correlation analysis, *exhibit* 721
- for simple regression analysis, 708, *exhibit* 707

SPSS-X, 375n.10, 623n.5
Srinivasan, V., 771n.21, 772n.23
Standard deviation, 411, 492–493, 496–497, 670
Standard error, 490
Standard error in regression analysis, 711
Standard error of the regression, 711
Standard industrial classification (SIC) system, 196n.1
Standardized regression coefficient, 713
Standard metropolitan statistical areas (SMSAs), 522–523, 547–548
Stanley, Julian C., 287, 295n.20, 302n.22, 325n.10, 328n.14, 331n.18
Stapel scale, 439–441
Starch INRA Hooper, Inc., 200
Starch Readership Reports, 200
Stasch, Stanley F., 172n.3, 586
Statistic, 485
Statistical designs, 334–338
Statistical efficiency, 514–516, 533, 534–535
Stevens, Joyce, 237n.33
Stewart, David W., 190n.14, 767n.16
Stocking, Carol, 378n.12
Stratified random sampling, 509–510
- defined, 510
- disproportionate, 520–522
- proportionate, 510–520
- simple cluster sampling vs., 512–514

Stratum charts, 831–832, *exhibit* 833
Straube, Linda D., 775n.28
Stress index, 769n.19
Structured, disguised questionnaire, 229
Structured, nondisguised questionnaire, 221
Structured observation, 246–247
Structured questions, 216–219, 367, 370–376
Suci, George J., 437
Sudman, Seymour, 145nn.8,10, 234n.24, 240n.38, 366n.2, 376n.11, 378n.12, 390n.19, 391n.21
Sullivan, Linda Varekamp, 205
Summers, Gene F., 407nn.1, 3, 414nn.7,8, 433n.27
Summers, John O., 312n.2
Superior Landscape Company, *case* 163–165
Survey of Buying Power, 203
Survey Sampling, Inc., 235
Sweeney, Dennis J., 489n.4, 492n.5, 701n.5, 702n.6, 716n.10
Syndicated data, 171–172
Syndicated service, 171–172, 176, 185–187
Syndicated sources of secondary data, 183–187, 198–203
Systematic error, 567
Systematic sampling, 529–533, *table* 535
Systems-oriented marketing research, 42

Tannenbaum, Percy H., 437
Tatham, Ronald L., 762n.9, 766n.14
Tauber, Edward M., 290n.12
Taylor, Thayer C., 42n.1, 46n.7, *exhibit* 47, 50n.11
Telephone surveys
- openers for, 399–401
- personal interviews and mail surveys vs., 229–240
- reducing nonresponse errors in, 578–582

Templeton, Jane Farley, 257n.12
Test market, simulated, 292–294
Test marketing, 289–290
Test-retest reliability, 443
Test statistic, defined, 650
Texize Chemicals, 271
Thematic apperception test (TAT), 225–226
Three-way table, *table* 635, 636
Thurstone, L. L., 433
Thurstone scale, 433
***Time* magazine survey,** 566
Timm, Paul R., 818n.6, 819n.7, 841n.16
Torgerson, Warren S., 407n.2
Total error, 564
Total sum of squares, 708, 733, 739
Trade associations, as sources of secondary data, 187–189
Trade-off approach in conjoint analysis, 772, 773
Transformed variables, 624–625
Traylor, Mark, 411n.5
Treatments, in ANOVA, 737–738
True experimental designs, 317–328
- features of, 317–320
- four-group, six-study form of, 326–328
- two-group, after-only form of, 323–325
- two-group, before-after form of, 320–323

True panel, 144–146
t-test, 664–665, 669–674, 698, 712, 719–720
Tuchfarber, Alfred J., 382n.15
Tullar, William L., 236n.31
Twedt, Dik Warren, 16n.14
Two-factors-at-a-time approach, in conjoint analysis, 772, 773
Two-group, after-only design, 323–325
Two-group, before-after design, 320–323
Two-group, *ex post facto* design, 315–316
Two-stage area sampling, 526–527
Two-tailed hypothesis test, 653–654
Two-way tabulations, 631–638
- graphical representation of, 835–839
- precautions in interpreting, 634–638

Tybjee, Tyzoon T., 235nn.25,28
Tybout, Alice M., 25n.38
Type I and Type II errors, 645–647

Unbalanced scale, 425
Unbiased estimate, 566
Unique zero point, 408
U.S. Census Bureau, 170, 171, 173–176, 177, 179, 476, 523
- publications of, 195–196

Univariate analysis, 655–677
- chi-square goodness-of-fit test, 656–660, 663–664
- defined, 655
- Kolmogorov-Smirnov test, *table* 656, 660–662, 663–664
- test for a single mean, *table* 656, 662–665
- test for a single proportion, *table* 657, 665–667
- test of two means (dependent samples), *table* 657, 672–674
- test of two means (independent samples), *table* 657, 667–672
- test of two proportions, *table* 657, 674–677
- types of tests in, *table* 656–657

Universe, 474
Urban, Glen L., 292n.17, 305n.24
Utility value, 772

Vadhia, Dinesh B., 46n.6, 49n.8, 52n.14
Validity
- threats to, in experiments, 295–306
- types of, in experiments, 287

Validity (*Continued*)
  types of, in measurement, 441–443
Variables
  behavioral, 414, 415
  classes of, 414–417
  dependent, 704, 714–716
  dummy, 779–783
  independent, 704, 714–716
  raw, 624
  recoded, 624–625
  transformed, 624–625
Variable transformation, 624–625
Venkatesan, M., 299n.21, 318n.5
Verille, Perry, 476n.3
Voice pitch analysis (VOPAN), 244, 420
Vovovich, B. C., 64n.2, 68n.11

Walker Research, Inc., *case* 78–80, 476, 585–586
Wallis, Lou, 42n.1
Walsh, Doris L., 176n.8, 767n.13
Ward, James C., 582n.11
Warner, W. Gregory, 411n.4, 426n.17
Warner-Lambert Company, 556, 557n
Warwick, Donald P., 373n.8, 382n.15
Waterman, Robert H., Jr., 210
Watson, Hugh J., 50n.13
Wayne Engineering Corporation, 101–109, 223
Webb, Eugene J., 216n.4, 246n.45
Webber, John, 42
Weighted average payoff, 114
Weinberg, Charles B., 815n.1
Weiss, Neil, 669n.6, *tables* 868–873
Wells, William D., 764n.12
Werner, D. Michael, 45n.5
Westfall, Ralph, 172n.3, 586
Whalen, Bernie, 190n.16, 243n.41, 246n.46
Wholesale data, sources of, 202–203
Wilcox, James B., 25n.38, 396n.25, 397n.26
Wildt, Albert R., 426n.18
Wilkens, P. L., 38n.5
Wilkie, William I., 418n.13
Williams, Thomas A., 489n.4, 492n.5, 701n.5, 702n.6, 716n.10
Wilson, H. W., Company, 204
Wilson, Jay F., 266n.27, 266n.29
Wind, Yoram, 139n.4, 330n.15, 771nn.20,22
Winona, Inc., 263n.26
Wise, Deborah C., 213n.2
Within sum of squares, 741
Wittink, Dick R., 775n.27
Wonnacott, Ronald J., 709n.8, 714n.9
Wonnacott, Thomas H., 709n.8, 714n.9
Word association test, 224
Worthing, Parker M., 292n.18
Worthington, Stan, 102, 103, 104, 108, 109
Wright, Deborah G., 64n.1, 66n.6, 68nn.8,10, 69n.12, 817n.1
Wrightsman, Lawrence S., 130n.2, 219n.5, 221n.6, 442n.35, 443n.36
Written reports
  desirable features of, 817–829
  graphical illustrations in, 829–840
  preparing, 816–829
  sections of, 819–823
WTTW-Chicago, *case* 344–348

Yakes, Nancy, 188n.13
Yankelovich Clancy Shulman, 178, *exhibit* 179, 201, 241–242, 292–294
Yankelovich, Skelly and White, 566
Yea-saying, 382
Y-intercept in regression analysis, 709
Young, Elliot, 243n.41
Young, Forrest W., 411n.5

Zale Corporation, 50
Zaltman, Gerald, 25n.38
Zeisel, Hans, 139, 306n.26, 616n.3, 633n.10
Zeisset, Paul T., *table* 175, 183n.10, *exhibit* 184
Zissman, Lorin, 759n.7
z-test, 664–665, 666–669, 675–677

LAPIN YLIOPISTO KIRJASTO
POISTETTU